

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 23, 2003, 10:43:52 ; Search time 289.144 Seconds
(without alignments)
10338.975 Million cell updates/sec

Title: US-09-717-321A-1

Perfect score: 123

Sequence: 1 caatgaaaaagttgttc.....ctcacagacccaaggtacc 123

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estov:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_esti:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_phg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	122	99.2	277	10	BF420446
2	122	99.2	292	9	AI598992
3	122	99.2	316	9	AI412434
4	122	99.2	404	10	BF549054

5	122	99.2	405	10	BE104696
6	122	99.2	407	9	AI009011
7	122	99.2	429	9	AI059212
8	122	99.2	455	9	AI012196
9	122	99.2	459	13	BQ780699
10	122	99.2	463	10	BF548903
11	122	99.2	485	9	AA899757
12	122	99.2	472	13	BQ190077
13	122	99.2	477	9	AW520555
14	122	99.2	489	9	AA859879
15	122	99.2	493	10	BF554834
16	122	99.2	522	14	C06842
17	122	99.2	565	12	BI273986
18	122	99.2	615	10	BG380448
19	122	99.2	639	10	BE111691
20	122	99.2	640	13	BQ780117
21	122	99.2	644	14	CA339407
22	122	99.2	687	12	BM389059
23	122	99.2	705	13	BQ200122
24	122	99.2	760	13	BQ191985
25	120.4	97.9	316	9	AI409795
26	120.4	97.9	557	10	BG665005
27	118.8	96.6	412	13	BO780657
28	118.8	96.6	517	9	AA859940
29	118.8	96.6	553	9	AA799542
30	118.8	96.6	636	13	BQ780658
31	118.8	96.6	638	12	BI283790
32	117.2	95.3	294	9	AI236740
33	117.2	95.3	469	12	BI276611
34	115.6	94.0	431	9	AI410870
35	115.6	94.0	446	12	BI286743
36	112.4	91.4	256	9	AI234716
37	110.8	90.1	427	9	AI178082
38	107.6	87.5	221	12	BM200395
39	107.6	87.5	225	10	BE136269
40	107.6	87.5	240	9	AW323013
41	107.6	87.5	269	12	BM200390
42	107.6	87.5	279	13	BU757687
43	107.6	87.5	283	9	AW122670
44	107.6	87.5	288	9	AV128839
45	107.6	87.5	294	14	W13019

ALIGNMENTS

RESULT 1

BF420446

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

BF420446 277 bp mRNA linear EST 28-NOV-2000
UI-R-BJ2-bpw-e-11-0-UI.s1 UI-R-BJ2 Rattus norvegicus cdna clone
UI-R-BJ2-bpw-e-11-0-UI.3', mRNA sequence.
GI:11408475
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 277)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
8889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu

Best Local Similarity 100.0%; Pred. No. 7.5e-31;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATTGAAAGTTTCTTAGTGGTCGAAAGCCCAACTGTGTTCGCCAGTGAG 60
|||||
Db 97 CAATTGAAAGTTTCTTAGTGGTCGAAAGCCCAACTGTGTTCGCCAGTGAG 156
|||||
QY 61 TTAGTTGTACAGAACGGCGTTAGCACTAGCGCTTGACAGAACCTCACAGACCCAAAGGT 120
|||||
Db 157 TTAGTTGTACAGAACGGCGTTAGCACTAGCGCTTGACAGAACCTCACAGACCCAAAGGT 216
|||||
QY 121 AC 122
||
Db 217 AC 218

RESULT 4
BF549054/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Rattus norvegicus (Norway rat)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 404)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
8889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
cDNA Library Preparation: M.B. Soares Lab Clone distribution:
clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward

FEATURES
source
1..404
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-A0-ag-b-01-0-UI"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-A0"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; This library
consists of a mixture of individually tagged normalized
libraries constructed from rat placenta, adult lung, brain
liver, kidney, heart, spleen, ovary, and muscle. The tag
is a string of 3-5 nucleotides present between the Not I
site and the oligo-dT track which allows identification of
the library of origin of a clone within the mixture."
92 a 98 c 83 g 131 t

BASE COUNT
ORIGIN
Query Match 99.2%; Score 122; DB 10; Length 404;
Best Local Similarity 100.0%; Pred. No. 8.3e-31;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAATTGAAAGTTTCTTAGTGGTCGAAAGCCCAACTGTGTTCGCCAGTGAG 60
|||||

Db 368 CAATTGAAAGTTTCTTAGTGGTCGAAAGCCCAACTGTGTTCGCCAGTGAG 309
QY 61 TTAGTTGTACAGAACGGCGTTAGCACTAGCGCTTGACAGAACCTCACAGACCCAAAGGT 120
|||||
Db 308 TTAGTTGTACAGAACGGCGTTAGCACTAGCGCTTGACAGAACCTCACAGACCCAAAGGT 249
|||||
QY 121 AC 122
||
Db 248 AC 247

RESULT 5
BE104696
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Rattus norvegicus (Norway rat)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 405)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
8889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
clones will be available through Research Genetics (www.resgen.com)
The following repetitive elements were found in this cDNA sequence:
1-22, >AT rich#Low complexity
Seq primer: M13 Forward
POLYA=Yes.

FEATURES
source
1..405
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BX0-ars-e-01-0-UI"
/dev_stage="embryonic 13 dpc"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-BX0"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BX0
library is derived from 13 dpc whole embryo tissue. For a
detailed description of the library from which this clone
was derived, please visit our web site at
rategest.eng.uiowa.edu.
TAG_SEQ=None found"
131 a 76 c 87 g 110 t 1 others

BASE COUNT
ORIGIN
Query Match 99.2%; Score 122; DB 10; Length 405;
Best Local Similarity 100.0%; Pred. No. 8.3e-31;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAATTGAAAGTTTCTTAGTGGTCGAAAGCCCAACTGTGTTCGCCAGTGAG 60
|||||

```

Db      117 CAATTGAAAAAGTTGTTCTAGTGTGCGAAGGCCCAACACTGTGTTCTTGCAGTGAG 176
QY      61 TTAGTTGTACAGAACGGCGTTAGCACTAGCGCTTGACAGAACCTCAGACACCCAAAGGT 120
Db      177 TTAGTTGTACAGAACGGCGTTAGCACTAGCGCTTGACAGAACCTCAGACACCCAAAGGT 236
QY      121 AC 122
Db      237 AC 238

RESULT 6
AI009011
LOCUS
DEFINITION EST203462 Normalized rat embryo, Bento Soares Rattus sp. cDNA clone
ACCESSION AI009011
VERSION    AI009011.1 GI:3222843
KEYWORDS
SOURCE
ORGANISM   Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE  1 (bases 1 to 407)
AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
TITLE      Rat Genome Project: Generation of a Rat EST (REBT) Catalog & Rat
            Gene Index
JOURNAL    Unpublished
COMMENT    Contact: Lee, NH
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13-21.
FEATURES   Location/Qualifiers
            1..407
             /organism="Rattus sp."
             /mol_type="mRNA"
             /db_xref="ATCC (inhost):2020087"
             /db_xref="taxon:10118"
             /clone="REMBH20"
             /dev_stage="embryo 8, 12, 18 dpc"
             /clone_lib="Normalized rat embryo, Bento Soares"
             /note="Vector: pT73Pac; Site_1: EcoRI; Site_2: NotI"
BASE COUNT 135 a 79 c 97 g 96 t
ORIGIN
Query Match 99.2%; Score 122; DB 9; Length 407;
Best Local Similarity 100.0%; Pred. No. 8.3e-31;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAATTGAAAAAGTTGTTCTAGTGTGCGAAGGCCCAACACTGTGTTCTTGCAGTGAG 60
Db 91 CAATTGAAAAAGTTGTTCTAGTGTGCGAAGGCCCAACACTGTGTTCTTGCAGTGAG 150
QY 61 TTAGTTGTACAGAACGGCGTTAGCACTAGCGCTTGACAGAACCTCAGACACCCAAAGGT 120
Db 151 TTAGTTGTACAGAACGGCGTTAGCACTAGCGCTTGACAGAACCTCAGACACCCAAAGGT 210
QY 121 AC 122
Db 211 AC 212

RESULT 7
AI059212
LOCUS
DEFINITION UI-R-Cl-1b-e-11-0-UI-s1 UI-R-Cl Rattus norvegicus cDNA clone
ACCESSION AI059212
            UI-R-Cl-1b-e-11-0-UI 3', mRNA sequence.
            AI059212

```

```

VERSION    AI059212.1 GI:3332989
KEYWORDS   EST.
SOURCE     Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE  1 (bases 1 to 429)
AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE      Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL    Genome Res. 6 (9), 791-806 (1996)
MEDLINE    97044477
PUBMED     8889548
COMMENT    Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            The sequence tag present in the cDNA between the NotI site and the
            oligo-dr track served to identify it as a clone from the normalized
            adult lung library. cDNA Library Preparation: M. Fatima Bonaldo,
            Ph.D. Clone distribution: clones will be available through Research
            Genetics
            Seq primer: M13 Forward.
FEATURES   Location/Qualifiers
            1..429
             /organism="Rattus norvegicus"
             /mol_type="mRNA"
             /strain="Sprague-Dawley"
             /db_xref="taxon:10116"
             /clone="UI-R-Cl-1b-e-11-0-UI"
             /dev_stage="Adult"
             /lab_host="DH10B (Life Technologies)"
             /clone_lib="UI-R-Cl"
             /note="Vector: pT73P-Pac (Pharmacia) with a modified
            polylinker; Site1: Not I; Site 2: Eco RI; The UI-R-Cl
            library is a subtracted library derived from the UI-R-Cl
            library, which is a subtracted library derived from the
            UI-R-A1 and UI-R-E1 libraries. The UI-R-A1 library
            consisted of a mixture of individually tagged normalized
            libraries constructed from rat placenta, adult lung,
            brain, liver, kidney, heart, spleen, ovary, and muscle.
            The UI-R-E1 library consisted of a mixture of
            individually tagged normalized libraries constructed from
            8, 12 and 18-day embryo. The tag is a string of 3-5
            nucleotides present between the Not I site and the
            oligo-dr track which allows identification of the library
            of origin of a clone within the mixture. The subtracted
            library (UI-R-Cl) was constructed as follows: PCR
            amplified cDNA inserts from UI-R-Cl clones from which 3'
            ESTs had been derived was used as a driver in a
            hybridization with the UI-R-Cl library in the form of
            single-stranded circles. The remaining single-stranded
            circles (subtracted library) was purified by
            hydroxyapatite column chromatography, converted to
            double-stranded cDNA and electroporated into DH10B
            bacteria (Life Technologies) to generate the UI-R-Cl
            library. This procedure has been previously described
            (Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
            1996)."
BASE COUNT 130 a 87 c 91 g 121 t
ORIGIN
Query Match 99.2%; Score 122; DB 9; Length 429;
Best Local Similarity 100.0%; Pred. No. 8.5e-31;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAATTGAAAAAGTTGTTCTAGTGTGCGAAGGCCCAACACTGTGTTCTTGCAGTGAG 60
Db 146 CAATTGAAAAAGTTGTTCTAGTGTGCGAAGGCCCAACACTGTGTTCTTGCAGTGAG 205

```



```

Qy 61 TTAGTTGTACAGAACGGCGTTAGCACTAGCGCTTGACAGACCTTCACAGACCCCAAGGT 120
    |||||||
Db 206 TTAGTTGTACAGAACGGCGTTAGCACTAGCGCTTGACAGACCTTCACAGACCCCAAGGT 265
    |||||||

Qy 121 AC 122
    ||
Db 266 AC 267

RESULT 8
LOCUS AI012196
DEFINITION EST206647 Normalized rat placenta, Bento Soares Rattus sp. cDNA
ACCESSION AI012196
VERSION AI012196.1 GI:3226028
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 455)
Lee, N.H., Glodek, A., Chandra, I., Mason, T.M., Quackenbush, J.,
Kerlavage, A.R. and Adams, M.D.
Rat Genome Project: Generation of a Rat EST (RST) Catalog & Rat
Gene Index
Unpublished
Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.

FEATURES
    source
    1..455
        /organism="Rattus sp."
        /mol_type="mRNA"
        /db_xref="taxon:10118"
        /clone="RPLAT85"
        /clone_lib="Normalized rat placenta, Bento Soares"
        /notes="Organ: placenta; Vector: pT7T3Pac; Site_1: EcoRI;
        Site_2: NotI"
BASE COUNT 156 a 92 c 102 g 105 t
ORIGIN
Query Match 99.2%; Score 122; DB 9; Length 455;
Best Local Similarity 100.0%; Pred. No. 8.8e-31;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAATTGAAAAAGTTTGTCTAGTGTGTCGAAAGGCCCAACACTGTCTTCCAGTGAG 60
    |||||||
Db 91 CAATTGAAAAAGTTTGTCTAGTGTGTCGAAAGGCCCAACACTGTCTTCCAGTGAG 150
    |||||||

Qy 61 TTAGTTGTACAGAACGGCGTTAGCACTAGCGCTTGACAGACCTTCACAGACCCCAAGGT 120
    |||||||
Db 151 TTAGTTGTACAGAACGGCGTTAGCACTAGCGCTTGACAGACCTTCACAGACCCCAAGGT 210
    |||||||

Qy 121 AC 122
    ||
Db 211 AC 212

RESULT 9
LOCUS BQ780699
DEFINITION UI-R-FF0-cpc-k-22-0-UI.s1 UI-R-FF0 Rattus norvegicus cDNA clone
ACCESSION BQ780699
VERSION BQ780699.1 GI:21989171
KEYWORDS EST.
459 bp mRNA linear EST 26-JUL-2002
UI-R-FF0-cpc-k-22-0-UI.s1 UI-R-FF0 Rattus norvegicus cDNA clone
UI-R-FF0-cpc-k-22-0-UI 3', mRNA sequence.

```

```

SOURCE
ORGANISM Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 459)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Jeff Stevens
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: DISTRIBUTION: Researchers may obtain clones
from Research Genetics (www.resgen.com).
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES
    Location/Qualifiers
    source
    1..459
        /organism="Rattus norvegicus"
        /mol_type="mRNA"
        /db_xref="taxon:10116"
        /clone="UI-R-FF0-cpc-k-22-0-UI"
        /tissue_type="Mixed tissues"
        /dev_stage="Adult"
        /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
        /clone_lib="UI-R-FF0"
        /note="Vector: pT7T3-Pac (Pharmacia) with a modified
        polylinker; Site 1: EcoR I; Site 2: Not I; UI-R-FF0 is a
        subtracted cDNA library containing the following tissue(s)
        ): Normal cartilage and SR-JWS tumor line. The
        subtraction was made according to Bonaldo, Lennon and
        Soares, Genome Research, 6:791-806, 1996. The
        oligonucleotide used to prime the synthesis of
        first-strand cDNA contains a library tag sequence that is
        located between the Not I site and the (dT)18 tail. The
        sequence tags for these libraries are: CTAATGGACG,
        CAATCTTGTA.
        TAG_LIB=UI-R-FF0
        TAG_TISSUE=cartilage
        TAG_SEQ=CTAATGGACG"
BASE COUNT 145 a 91 c 103 g 120 t
ORIGIN
Query Match 99.2%; Score 122; DB 13; Length 459;
Best Local Similarity 100.0%; Pred. No. 8.8e-31;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAATTGAAAAAGTTTGTCTAGTGTGTCGAAAGGCCCAACACTGTCTTCCAGTGAG 60
    |||||||
Db 114 CAATTGAAAAAGTTTGTCTAGTGTGTCGAAAGGCCCAACACTGTCTTCCAGTGAG 173
    |||||||

Qy 61 TTAGTTGTACAGAACGGCGTTAGCACTAGCGCTTGACAGACCTTCACAGACCCCAAGGT 120
    |||||||
Db 174 TTAGTTGTACAGAACGGCGTTAGCACTAGCGCTTGACAGACCTTCACAGACCCCAAGGT 233
    |||||||

Qy 121 AC 122
    ||
Db 234 AC 235

RESULT 10
BF548903/c

```

LOCUS BF548903 463 bp mRNA linear EST 11-DEC-2000
DEFINITION UI-R-A0-ag-b-08-0-UI.r1 UI-R-A0 Rattus norvegicus cDNA clone
UI-R-A0-ag-b-08-0-UI 5', mRNA sequence.
ACCESSION BF548903
VERSION BF548903.1 GI:11640010
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
REFERENCE 1 (bases 1 to 463)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLES Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PubMed 8889548
COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
cDNA Library Preparation: M.B. Soares Lab Clone distribution:
clones will be available through Research Genetics (www.resgen.com)
This clone is also available through the I.M.A.G.E. Consortium at
LLNL (info@image.llnl.gov). IMAGE ID= 1791813
Seq primer: M13 Forward.
Location/Qualifiers
1..463
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-A0-ag-b-08-0-UI"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-A0"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; This library consists of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, and muscle. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture."
BASE COUNT 109 a 106 c 93 g 155 t
ORIGIN
Query Match 99.2%; Score 122; DB 10; Length 463;
Best Local Similarity 100.0%; Pred. No. 8.8e-31;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAATTGAAAGTTTGTCTAGTGTGCGAAGGCCCAACACTGTCTTTCGCACTGAG 60
DB 368 CAAATGAAAGTTTGTCTAGTGTGCGAAGGCCCAACACTGTCTTTCGCACTGAG 309
QY 61 TTAGGTTGTACAGACGGCGTTAGCACTAGCGCTTGACAGAACTCAGACACCAAGGT 120
DB 308 TTAGGTTGTACAGACGGCGTTAGCACTAGCGCTTGACAGAACTCAGACACCAAGGT 249
QY 121 AC 122
DB 248 AC 247
RESULT 11
AA899757 465 bp mRNA linear EST 05-FEB-1999
LOCUS UI-R-E0-da-b-11-0-UI.s1 UI-R-E0 Rattus norvegicus cDNA clone
DEFINITION UI-R-E0-da-b-11-0-UI 3' similar to gi|464185|dbj|D25274|HMP02ST9

ACCESSION AA899757
VERSION AA899757.1 GI:4232251
KEYWORDS EST.
SOURCE Homo sapiens mRNA, clone:PO2ST9, mRNA sequence.
ORGANISM Rattus norvegicus (Norway rat)
REFERENCE 1 (bases 1 to 465)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLES Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PubMed 8889548
COMMENT On Apr 7, 1998 this sequence version replaced gi:3035111.
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics
Seq primer: M13 Forward.
Location/Qualifiers
1..465
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-E0-da-b-11-0-UI"
/dev_stage="embryonic"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-E0"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: NotI; Site 2: EcoRI; This library consists of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture."
BASE COUNT 155 a 93 c 99 g 118 t
ORIGIN
Query Match 99.2%; Score 122; DB 9; Length 465;
Best Local Similarity 100.0%; Pred. No. 8.8e-31;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAATTGAAAGTTTGTCTAGTGTGCGAAGGCCCAACACTGTCTTTCGCACTGAG 60
DB 112 CAAATGAAAGTTTGTCTAGTGTGCGAAGGCCCAACACTGTCTTTCGCACTGAG 171
QY 61 TTAGGTTGTACAGACGGCGTTAGCACTAGCGCTTGACAGAACTCAGACACCAAGGT 120
DB 172 TTAGGTTGTACAGACGGCGTTAGCACTAGCGCTTGACAGAACTCAGACACCAAGGT 231
QY 121 AC 122
DB 232 AC 233
RESULT 12
BQ190077 472 bp mRNA linear EST 30-APR-2002
LOCUS UI-R-CN1-cjs-h-07-0-UI.s3 UI-R-CN1 Rattus norvegicus cDNA clone
DEFINITION UI-R-CN1-cjs-h-07-0-UI 3', mRNA sequence.
ACCESSION BQ190077
VERSION BQ190077.1 GI:20365628
KEYWORDS EST.

SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 472)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
8889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized rat eye library cDNA Library Preparation: M.B. Soares
Lab Clone distribution: clones will be available through Research
Genetics (www.regen.com)
Seq primer: M13 Forward
POLYA=yes.

FEATURES
source
1. .472
Location/Qualifiers
/organism="Rattus norvegicus"
/mol_type="RNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CNI-G1s-h-07-0-UI"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-CNI"
/note="Vector: pTT3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-CNI
library is a subtracted library derived from the following
pool of seven normalized rat libraries: normalized rat
seminal vesicles, normalized rat penis, normalized rat
bladder, normalized rat cervix, normalized rat brown
adipose, normalized rat fundus, and normalized rat
salivary gland. It was constructed according to the
procedure described by Bonaldo, Lennon & Soares (Genome
Research Genome 6: 791-806, 1996). For construction of
the CNI library, plasmid DNA from the pool of normalized
libraries was electroporated into competent bacteria for
the production of single-stranded circular DNA. This was
then used as a tracer in a subtractive hybridization with
a driver (PCR amplified inserts from a plasmid DNA template
preparation) comprising: a) a pool of about 34,000 clones
from the Rat Unigene Set corresponding to plates R-5-AA-NN
excluding plates R-5-MM and MN. This pool represented 40%
of the final driver population. b) a pool of about 29,000
clones from subtracted libraries CA0 and CA1 corresponding
to plates R-CA0-AWV through R-CA0-AXS, R-CA0-AZX through
R-CA0-BAZ, R-CA0-BFE through R-CA0-BHY, R-CA0-BJS,
R-CA0-BKS, R-CA0-BKV-V, R-CA0-BKJ-K, R-CA0-BKP through
R-CA0-BMC through R-CA0-BME, R-CA0-BNS, R-CA0-BOB through
R-CA0-BOU, R-CA0-BPA through R-CA0-BPG, R-CA1-BBA through
R-CA1-BDA, R-CA1-BHZ through R-CA1-BJF, R-CA1-BJR,
R-CA1-BJT through R-CA1-BKB, R-CA1-BKD, R-CA1-BKF,
R-CA1-BKT, R-CA1-BKT, R-CA1-BLF, R-CA1-BLH through
R-CA1-BLN, R-CA1-BJS, R-CA1-BLU-V, R-CA1-BNR, and
R-CA1-BUE. The resulting pool represented 20% of the
final driver population. c) a pool of about 15,000 clones
from non-normalized libraries CS0s, CT0s, CW0s, CX0s
and normalized libraries CS0, CT0, CU0, CW0, and CX0

corresponding to plates R-CS0s-CBD through R-CS0s-CBO,
R-CT0s-CAM through R-CT0s-CAX, R-CU0s-CBP through
R-CU0s-CCA, R-CW0s-CCB through R-CW0s-CCM, R-CX0s-CCN
through R-CX0s-CCX, R-CS0-BSD, R-CS0-BTD through R-CS0-BTV
R-CS0-BVM, R-CT0-BTW through R-CT0-BUP, R-CT0-BVN,
R-CU0-BUQ through R-CU0-BVL, R-CW0-BVW through R-CW0-BWP,
R-CX0-BXN through R-CX0-BXO, R-CX0-BWQ through R-CX0-BXM.
The resulting pool represented 5% of the final driver
population. d) a pool of about 5,000 clones (1,000 from
non-normalized eye library CV0 and 4,000 from normalized
eye library CV1) corresponding to plates R-CV0-BRH through
R-CV0-BRR, R-CV1-BRS through R-CV1-BSC, R-CV1-BSE through
R-CV1-BtC, and R-CV1-BVO through R-CV1-BVU. This pool
represented about 5% of the final driver population. e) A
pool of about 10,000 clones from subtracted library BS2,
BV0 and BV0p (7-9.5 kb cDNA library fraction from rat
whole embryo), and BX0 (0.5-7kb cDNA library fraction from
rat whole embryo) corresponding to plates R-BS2-BDB
through R-BS2-BFB, R-BV0-ANK through R-BV0-ANR, R-BV0p-AOI
through R-BV0p-AOX, and R-BX0-AQY through R-BX0-ASH. The
resulting pool represented 5% of the final driver
population. f) a pool of about 7,000 clones from the
seven non-normalized libraries that make up the tracer
including CV0, CZ0, DA0, DB0, DC0, DD0, and DE0
corresponding to plates R-CY0-BXP through R-CY0-BXZ,
R-CZ0-BXA through R-CZ0-BYI, R-CZ0-BZB-C, R-DA0-BYJ
through R-DA0-BYP, R-DA0-BZD through R-DA0-BZH, R-DB0-BYQ
through R-DB0-BZA, R-DC0-BZI through R-DC0-BZQ, R-DC0-CAY
through R-DC0-CBA, R-DD0-BZR through R-DD0-CAA, The
R-DD0-CBB-C, and R-DE0-CAB through R-DE0-CAL. The
resulting pool represented about 10% of the final driver
population. g) a pool of about 2,000 clones from the pool
of normalized libraries, CNO, that makes up the tracer.
The corresponding plates are R-CNO-BKM through R-CNO-BLD,
R-CNO-BLG, R-CNO-BLP through R-CNO-BLR, R-CNO-BLT,
R-CNO-BLW-X, R-CNO-BMB, and R-CNO-BMF through R-CNO-BML.
This pool represented 5% of the final driver population.
h) a pool of the 28 most abundant clones in the CNO pool
corresponding to the following addresses: bkx-a-09-0-UI,
bkx-b-09-0-UI, bkx-b-11-0-UI, bkx-b-10-0-UI, bkx-d-01-0-UI,
bkx-d-06-0-UI, bkx-g-08-0-UI, bkx-h-12-0-UI,
bkx-a-05-0-UI, bkz-a-06-0-UI, bkz-a-11-0-UI, bkz-c-06-0-UI,
bkz-c-09-0-UI, bkz-d-10-0-UI, bla-a-01-0-UI,
bla-a-02-0-UI, bla-f-04-0-UI, bla-g-07-0-UI, bla-g-12-0-UI,
blb-a-12-0-UI, blb-f-02-0-UI, blc-a-11-0-UI,
blc-e-95-0-UI, bld-1-08-0-UI, bld-f-02-0-UI, blg-h-04-0-UI,
blr-a-05-0-UI, blt-f-08-0-UI. This pool represented 5%
of the final driver population. i) One abundant CNO clone
(corresponding to the address bkz-a-11-0-UI) was digested
with Not I and Eco RI and the resulting insert was gel
purified. This purified insert was added directly to the
driver so that it represented 5% of the final driver
population.

TAG LI8-UI-R-CNI
TAG-TISSUE-rat eye
TAG_SEQ=CAGCC"

BASE COUNT 154 a 102 g 123 t
ORIGIN

Query Match 99.2%; Score 122; DB 13; Length 472;
Best Local Similarity 100.0%; Pred.No. 8.9e-31;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAATTGAAAAAGTTTGTCTAGTGGTCCGAAGGCCCAACTGTGTCTTCCAGTGCAG 60
|||
Db 113 CAATTGAAAAAGTTTGTCTAGTGGTCCGAAGGCCCAACTGTGTCTTCCAGTGCAG 172
|||
QY 61 TTAGGTTTACAGAACGGCGTTAGCACTAGCGCTTGACAGAACCTTCACAGACCCCAAGGT 120
|||
Db 173 TTAGGTTTACAGAACGGCGTTAGCACTAGCGCTTGACAGAACCTTCACAGACCCCAAGGT 232
|||
QY 121 AC 122
|||

```

Db          233 AC 234
RESULT 13
AW520555
LOCUS
DEFINITION UI-R-BJ0p-afx-g-09-0-UI.s1 477 bp mRNA linear EST 06-MAR-2000
            UI-R-BJ0p-afx-g-09-0-UI 3', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 477)
Normalizaton and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
8889548
COMMENT
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized Av canal at 15 dpc library cDNA Library Preparation:
M.B. Soares Lab Clone distribution: clones will be available
through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES
source
Location/Qualifiers
1..477
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BJ0p-afx-g-09-0-UI"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-BJ0p"
/note="Vector: pTT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-BJ0p
library is a subtracted library derived from the UI-R-AAL,
UI-R-AB1, UI-R-AC1, UI-R-AD1, UI-R-AE1, UI-R-AF1, and
UI-R-AG1 libraries. These libraries represent tissues from
rat atrium at 16.5 dpc, ventricle at 16.5 dpc, AV canal
at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc, AV
canal at 15 dpc, and ventricle at 13 dpc. The tag is a
string of 5-6 nucleotides present between the Not I site
and the oligo-dT track. The library was constructed as
described by Bonaldo, Lennon and Soares, Genome Research
6: 791-806, 1996.
TAG_LIB=UI-R-BJ0p
TAG_TISSUE=AV canal at 15 dpc
TAG_SEQ=GAAGG"
BASE COUNT 157 a 103 g 124 t
ORIGIN
Query Match 99.2%; Score 122; DB 9; Length 477;
Best Local Similarity 100.0%; Pred. No. 8.9e-31;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db          233 AC 234
RESULT 14
AW520555
LOCUS
DEFINITION UI-R-E0-cc-c-06-0-UI.s1 489 bp mRNA linear EST 03-JUL-1999
            UI-R-E0-cc-c-06-0-UI 3', similar to dbj|D25274|HUMPO2St9 Human
            randomly sequenced mRNA, mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 489)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalizaton and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
8889548
COMMENT
On Mar 10, 1998 this sequence version replaced gi:2949399.
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
The sequence tag present in the cDNA between the NotI site and the
oligo-dT track served to identify it as a clone from the normalized
adult 18-Day-Embryo library. cDNA Library Preparation: M. Fatima
Bonaldo, Ph.D. Clone distribution: clones will be available through
Research Genetics This clone is also available through the
I.M.A.G.E. Consortium at LLNL (info@image.llnl.gov). IMAGE
ID=1770064
Seq primer: M13 Forward
POLYA=No.

FEATURES
source
Location/Qualifiers
1..489
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-E0-cc-c-06-0-UI"
/dev_stage="embryonic"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-E0"
/note="Vector: pTT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: NotI; Site 2: EcoRI; This library
consists of a mixture of individually tagged normalized
libraries constructed from 8, 12 and 18-day embryo. The
tag is a string of 3-5 nucleotides present between the
Not I site and the oligo-dT track which allows
identification of the library of origin of a clone within
the mixture."
BASE COUNT 158 a 96 c 107 g 128 t
ORIGIN
Query Match 99.2%; Score 122; DB 9; Length 489;
Best Local Similarity 100.0%; Pred. No. 9e-31;

```

Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATTGAAAAAGTTTGTCTAGTGTGTCGAAAGGCCAACACACTGTGTTCTTCCAGTGAG 60
|||
Db 112 CAATTGAAAAAGTTTGTCTAGTGTGTCGAAAGGCCAACACACTGTGTTCTTCCAGTGAG 171
|||
QY 61 TTAGGTTGTACAGAACGGCGTTAGCACTAGCGCTTGACAGAACCTTCACAGACCCCAAGGT 120
|||
Db 172 TTAGGTTGTACAGAACGGCGTTAGCACTAGCGCTTGACAGAACCTTCACAGACCCCAAGGT 231
|||
QY 121 AC 122
||
Db 232 AC 233

RESULT 15
BF554834/c
LOCUS
DEFINITION UI-R-E0-cc-c-06-0-UI.r1 UI-R-E0 Rattus norvegicus cDNA clone
ACCESSION BF554834
VERSION BF554834.1 GI:11664606
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
REFERENCE 1 (bases 1 to 493)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
cDNA Library Preparation: M.B. Soares Lab Clone distribution:
clones will be available through Research Genetics (www.resgen.com)
This clone is also available through the I.M.A.G.E. Consortium at
LLNL (info@image.llnl.gov). IMAGE ID= 1770064 The following
repetitive elements were found in this cDNA sequence: 463-485,
>AT rich#Low complexity
Seq primer: M13 Forward.

FEATURES
source
1..493
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-E0-cc-c-06-0-UI"
/dev_stage="embryonic"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-E0"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: NotI; Site 2: EcoRI; This library
consists of a mixture of individually tagged normalized
libraries constructed from 8, 12 and 18-day embryo. The
tag is a string of 3-5 nucleotides present between the
Not I site and the oligo-dT track which allows
identification of the library of origin of a clone within
the mixture."
BASE COUNT 130 a 109 c 95 g 159 t
ORIGIN

Query Match 99.2%; Score 122; DB 10; Length 493;
Best Local Similarity 100.0%; Pred.No. 9.1e-31;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATTGAAAAAGTTTGTCTAGTGTGTCGAAAGGCCAACACACTGTGTTCTTCCAGTGAG 60
|||
Db 368 CAATTGAAAAAGTTTGTCTAGTGTGTCGAAAGGCCAACACACTGTGTTCTTCCAGTGAG 309
|||
QY 61 TTAGGTTGTACAGAACGGCGTTAGCACTAGCGCTTGACAGAACCTTCACAGACCCCAAGGT 120
|||
Db 308 TTAGGTTGTACAGAACGGCGTTAGCACTAGCGCTTGACAGAACCTTCACAGACCCCAAGGT 249
|||
QY 121 AC 122
||
Db 248 AC 247

Search completed: November 23, 2003, 15:49:53
Job time : 297.144 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 23, 2003, 09:17:01 ; Search time 35.0599 Seconds
(without alignments)
9470.385 Million cell updates/sec

Title: US-09-717-321A-1

Perfect score: 123

Sequence: 1 caatgaaaaagtgtgttc.....ctcacagaccccaaggtaacc 123

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq 19Jun03.*

```

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	123	100.0	123	22	AAH22395
2	123	100.0	348	22	AAH22399
3	123	100.0	1266	22	AAH22396
4	106.6	86.7	422	21	AAH89694
5	96.6	78.5	447	21	AAH89693
6	79.6	64.7	352	22	AAH22400
7	79.6	64.7	1232	24	ABK83623
8	79.6	64.7	1232	24	ABN95134
					Human rac1 genomic
					Human rac1 gene re
					Human rac1 contig
					Mouse Rab2 nucleot
					Mouse Exol103 nucle
					Human rac1 gene re
					Human cDNA differe
					Gene #1632 used to

C 9	79.6	64.7	1318	23	AAH22395	DNA encoding novel
C 10	79.6	64.7	2051	21	AAH22396	lung cancer associ
C 11	79.6	64.7	28567	25	ABT17030	Human MP21 gene Ra
C 12	57	46.3	353	25	ABX46267	Bovine EST associa
C 13	38.2	31.1	3740	22	AAH07296	Human reproductiv
C 14	34	27.6	50	24	ABZ04062	Human leukocyte ge
C 15	31.4	25.5	605	21	AAH16478	Human colon cancer
C 16	31.4	25.5	605	21	ABU38065	Human colon tumour
C 17	28.8	23.4	202001	24	ABH52506	Human transporter
C 18	27.4	22.3	412	24	ABN94214	Gene #712 used to
C 19	27.4	22.3	412	24	ABL62818	Breast cancer rela
C 20	27.4	22.3	412	24	ABL63018	Breast cancer rela
C 21	27.4	22.3	412	24	ABL63234	Breast cancer rela
C 22	27.4	22.3	412	24	ABH80240	Human ovarian canc
C 23	27.4	22.3	1089	21	AAH05142	Arabidopsis thalia
C 24	27.4	22.3	1091	21	AAH05147	Arabidopsis thalia
C 25	27.4	22.3	1658	24	ABT10169	Human breast cance
C 26	27.4	22.3	1689	22	AAH59609	Human cell cycle a
C 27	27.2	22.1	9965	24	ABL33526	Human immune syste
C 28	27	22.0	363	24	ABT199532	Mouse ischaemic co
C 29	27	22.0	6693	25	ACA04014	cDNA downregulated
C 30	26.8	21.8	396	25	ABX05806	S. pneumoniae type
C 31	26.8	21.8	1617	24	ABK63792	Rat sequence diffe
C 32	26.8	21.8	6827	19	AAV52193	Streptococcus pneu
C 33	26.8	21.8	2162598	25	ABH56454	Streptococcus pneu
C 34	26.6	21.6	1030	21	AAC34296	Arabidopsis thalia
C 35	26.6	21.6	81001	22	AAH30035	Human apolipoprote
C 36	26.4	21.5	279	24	ABL60574	Human monocyte che
C 37	26.4	21.5	510	25	ABX74789	Human cDNA sequenc
C 38	26.4	21.5	1035	21	AAH51499	Human growth facto
C 39	26.4	21.5	1035	24	ABH68613	DNA representing a
C 40	26.4	21.5	1035	25	ABX93147	Degenerate DNA enc
C 41	26.4	21.5	534720	19	AAV30458	Rhizobium species
C 42	26.4	21.5	536165	19	AAV30459	Rhizobium species
C 43	26.2	21.3	903	24	ABN70636	Streptococcus poly
C 44	26.2	21.3	2091	18	AAV25107	H. pylori cell env
C 45	26.2	21.3	2091	20	AAH75826	H. pylori outer me

ALIGNMENTS

RESULT 1
AAH22395
ID AAH22395 standard; DNA; 123 BP.
XX AC
XX AAH22395;
XX DT 22-AUG-2001 (first entry)
XX DE Human rac1 genomic fragment probable 3' UTR SEQ ID NO:1.
XX DE Identification; toxic; hepatotoxic; differential gene expression;
KW NSAIID; non-steroidal antiinflammatory drug; ds.
XX OS Homo sapiens.
XX OS
XX WO200138579-A2.
XX PD 31-MAY-2001.
XX PF 21-NOV-2000; 2000WO-US32049.
XX PR 22-NOV-1999; 99US-0166923.
PR 18-FEB-2000; 2000US-0183531.
PR 20-NOV-2000; 2000US-0717321.
XX (CURA-) CURAGEN CORP.
PI Gould-Rothberg BE, Dipippo VA, Ramseh TM, Gerwein RW;
DR WPI; 2001-355948/37.

PT Screening hepatotoxic agent comprises contacting test cell population
 PT expressing RISKMARKER or INJURYMARKER with agent, comparing expression
 PT with reference population and identifying difference in expression
 PT levels

XX Disclosure; Page 7; 76pp; English.

XX The present invention describes a method of screening a test agent for
 CC hepatotoxicity. The method comprises: (a) providing a test cell
 CC population comprising a cell capable of expressing one or more nucleic
 CC acid sequences selected from the group consisting of RISKMARKER 1-8
 CC and INJURYMARKER 1-10; (b) contacting the test cell population with a
 CC test agent; (c) measuring expression of one or more of the nucleic
 CC acid sequences in the test cell population; (d) comparing the
 CC expression of the nucleic acid sequence in the test cell population to
 CC the expression of the nucleic acid sequence in an reference cell
 CC population comprising at least one cell whose exposure status to a
 CC hepatotoxic agent in known; and (e) identifying a difference in
 CC expression levels of the RISKMARKER or INJURYMARKER sequences, if
 CC present, in the test cell population and reference cell population.
 CC The method is useful for identifying a hepatotoxic agent. The present
 CC sequence is given in the exemplification of the present invention.

XX SQ Sequence 123 BP; 36 A; 28 C; 30 G; 29 T; 0 other;

Query Match 100.0%; Score 123; DB 22; Length 123;
 Best Local Similarity 100.0%; Pred. No. 1.28-35;
 Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATTGAAAAAGTTTGTCTAGTGTGCGAAGGCCCAACACTGTGTTCTTCCAGTGAG 60

Db 1 CAATTGAAAAAGTTTGTCTAGTGTGCGAAGGCCCAACACTGTGTTCTTCCAGTGAG 60

QY 61 TTAGTTTGTACAGACGGCGTTAGCACTAGCGCTTGACAGAACCTCAGAGCCCAAGGT 120

Db 61 TTAGTTTGTACAGACGGCGTTAGCACTAGCGCTTGACAGAACCTCAGAGCCCAAGGT 120

QY 121 ACC 123

Db 121 ACC 123

RESULT 2

AAH22399/c

ID AAH22399 standard; DNA; 348 BP.

XX AC AAH22399;

DT 22-AUG-2001 (first entry)

DE Human rac1 gene related nucleotide sequence #3.

XX Identification; toxic; hepatotoxic; differential gene expression;

KW NSAID; non-steroidal antiinflammatory drug; ds.

XX OS Homo sapiens.

XX WO200138579-A2.

XX PD 31-MAY-2001.

XX PF 21-NOV-2000; 2000WO-US32049.

XX PR 22-NOV-1999; 99US-0166923.

XX PR 18-FEB-2000; 2000US-0183531.

XX PR 20-NOV-2000; 2000US-0717321.

XX PA (CURA-) CURAGEN CORP.

XX PI Gould-Rothberg BE, Dipippo VA, Ramseh TM, Gerwein RW;

XX DR WPI; 2001-355948/37.

XX XX

PT Screening hepatotoxic agent comprises contacting test cell population
 PT expressing RISKMARKER or INJURYMARKER with agent, comparing expression
 PT with reference population and identifying difference in expression
 PT levels

XX Disclosure; Page 8-9; 76pp; English.

XX The present invention describes a method of screening a test agent for
 CC hepatotoxicity. The method comprises: (a) providing a test cell
 CC population comprising a cell capable of expressing one or more nucleic
 CC acid sequences selected from the group consisting of RISKMARKER 1-8
 CC and INJURYMARKER 1-10; (b) contacting the test cell population with a
 CC test agent; (c) measuring expression of one or more of the nucleic
 CC acid sequences in the test cell population; (d) comparing the
 CC expression of the nucleic acid sequence in the test cell population to
 CC the expression of the nucleic acid sequence in an reference cell
 CC population comprising at least one cell whose exposure status to a
 CC hepatotoxic agent in known; and (e) identifying a difference in
 CC expression levels of the RISKMARKER or INJURYMARKER sequences, if
 CC present, in the test cell population and reference cell population.
 CC The method is useful for identifying a hepatotoxic agent. The present
 CC sequence is given in the exemplification of the present invention.

XX SQ Sequence 348 BP; 103 A; 76 C; 67 G; 102 T; 0 other;

Query Match 100.0%; Score 123; DB 22; Length 348;
 Best Local Similarity 100.0%; Pred. No. 1.78-35;
 Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATTGAAAAAGTTTGTCTAGTGTGCGAAGGCCCAACACTGTGTTCTTCCAGTGAG 60

Db 212 CAATTGAAAAAGTTTGTCTAGTGTGCGAAGGCCCAACACTGTGTTCTTCCAGTGAG 153

QY 61 TTAGTTTGTACAGACGGCGTTAGCACTAGCGCTTGACAGAACCTCAGAGCCCAAGGT 120

Db 152 TTAGTTTGTACAGACGGCGTTAGCACTAGCGCTTGACAGAACCTCAGAGCCCAAGGT 93

QY 121 ACC 123

Db 92 ACC 90

RESULT 3

AAH22396

ID AAH22396 standard; DNA; 1266 BP.

XX AC AAH22396;

DT 22-AUG-2001 (first entry)

DE Human rac1 contig SEQ ID NO:2.

XX Identification; toxic; hepatotoxic; differential gene expression;

KW NSAID; non-steroidal antiinflammatory drug; ds.

XX OS Homo sapiens.

XX WO200138579-A2.

XX PD 31-MAY-2001.

XX PF 21-NOV-2000; 2000WO-US32049.

XX PR 22-NOV-1999; 99US-0166923.

XX PR 18-FEB-2000; 2000US-0183531.

XX PR 20-NOV-2000; 2000US-0717321.

XX PA (CURA-) CURAGEN CORP.

XX PI Gould-Rothberg BE, Dipippo VA, Ramseh TM, Gerwein RW;

XX DR WPI; 2001-355948/37.

XX XX

PT Screening hepatotoxic agent comprises contacting test cell population
PT expressing RISKMARKER or INJURYMARKER with agent, comparing expression
PT with reference population and identifying difference in expression
PT levels
XX
PS Disclosure; Page 7; 76pp; English.
XX
XX The present invention describes a method of screening a test agent for
CC hepatotoxicity. The method comprises: (a) providing a test cell
CC population comprising a cell capable of expressing one or more nucleic
CC acid sequences selected from the group consisting of RISKMARKER 1-8
CC and INJURYMARKER 1-10; (b) contacting the test cell population with a
CC test agent; (c) measuring expression of one or more of the nucleic
CC acid sequences in the test cell population; (d) comparing the
CC expression of the nucleic acid sequence in the test cell population to
CC the expression of the nucleic acid sequence in a reference cell
CC population comprising at least one cell whose exposure status to a
CC hepatotoxic agent is known; and (e) identifying a difference in
CC expression levels of the RISKMARKER or INJURYMARKER sequences, if
CC present, in the test cell population and reference cell population.
CC The method is useful for identifying a hepatotoxic agent. The present
CC sequence is given in the exemplification of the present invention.
XX
SQ Sequence 1266 BP; 385 A; 258 C; 285 G; 338 T; 0 other;

Query Match 100.0%; Score 123; DB 25; Length 1266;
Best Local Similarity 100.0%; Pred. No. 2.8e-35;
Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATTGAAAAAGTTTGTCTAGTGGTCGAAAGGCCCAACACTGTGTTCCAGTGAG 60
Db 146 CAATTGAAAAAGTTTGTCTAGTGGTCGAAAGGCCCAACACTGTGTTCCAGTGAG 205

QY 61 TTAGTTGTACAGAACGGCGTTAGCTAGCGTTGACAGAACTCTACAGACCCAAAGGT 120
Db 206 TTAGTTGTACAGAACGGCGTTAGCTAGCGTTGACAGAACTCTACAGACCCAAAGGT 265

QY 121 ACC 123
Db 266 ACC 268

RESULT 4
AAA89694/c
ID AAA89694 standard; cDNA; 422 BP.
XX
AC AAA89694;
XX
XX 08-JAN-2001 (first entry)
XX
XX Mouse Rab2 nucleotide sequence #1.
XX
XX Exocytosis pathway protein; Exo protein; antiparkinsonian; antidiabetic;
KW anti-allergic; antiasthmatic; nootropic; neuroprotective; anticonvulsant;
KW vulnary; asthma; inflammation; allergy; Chediak-Higashi syndrome; CHS;
KW Alzheimer's disease; Parkinson's disease; Huntington's disease; diabetes;
XX digestion disorder; wound healing disorder; gene therapy; ss.
XX
OS Mus sp.
XX
XX WO200043419-A2.
XX
XX 27-JUL-2000.
XX
XX 20-JAN-2000; 2000WO-US01431.
XX
PR 20-JAN-1999; 99US-0116534.
PR 26-JAN-1999; 99US-0117274.
PR 26-JAN-1999; 99US-0117308.
PR 26-JAN-1999; 99US-0117309.
PR 26-JAN-1999; 99US-0117312.
PR 01-FEB-1999; 99US-0118177.
PR 01-FEB-1999; 99US-0118178.

PR 01-FEB-1999; 99US-0118179.
PR 09-FEB-1999; 99US-0119286.
PR 11-FEB-1999; 99US-0119998.
PR 11-FEB-1999; 99US-0119759.
XX
XX (RIGE-) RIGEL PHARM INC.
XX
XX Luo Y;
PI
XX WPI; 2000-482908/42.
DR
XX New nucleic acids encoding Exo proteins which are useful in the
PT diagnosis, treatment or prevention of exocytosis-mediated disorders
PT such as asthma, inflammation and allergies -
XX
XX Disclosure; Page 271-272; 305pp; English.
PS
XX The present sequence encodes a polypeptide which is associated with
CC the exocytosis pathway. cDNA molecules encoding proteins involved in
CC exocytosis have been isolated by yeast one-hybrid and two-hybrid
CC screening. Novel proteins, termed Exo proteins, have been identified that
CC interact with known exocytosis-associated proteins such as GS27, alpha
CC snap, unc18-1, vamp3, snap-23, and the rab family of proteins.
CC Exo proteins and their agonists and antagonists are useful in the
CC diagnosis, treatment or prevention of exocytosis-mediated disorders
CC such as asthma, inflammation, allergies, Chediak-Higashi Syndrome
CC (CHS), Alzheimer's disease, Parkinson's disease, Huntington's disease,
CC diabetes, digestion disorders and wound healing disorders.
CC The nucleic acids, antagonists or agonists of Exo proteins are useful
CC in gene therapy. The nucleic acids are also useful for generating
CC transgenic or knock-out animals which can be used in the
CC development and screening of therapeutically useful reagents.
XX
SQ Sequence 422 BP; 95 A; 92 C; 83 G; 148 T; 4 other;

Query Match 86.7%; Score 106.6; DB 21; Length 422;
Best Local Similarity 91.8%; Pred. No. 2.2e-29;
Matches 112; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 CAATTGAAAAAGTTTGTCTAGTGGTCGAAAGGCCCAACACTGTGTTCCAGTGAG 60
Db 378 CAATTGAAAAAGTTTGTCTAGTGGTTGAAGGCCCAACACTGTGTTCCAGTGAG 319

QY 61 TTAGTTGTACAGAACGGCGTTAGCTAGCGTTGACAGAACTCTACAGACCCAAAGGT 120
Db 318 TTAGTTGTACAGAACGGCGTTAGCTAGCGTTGACAGAACTCTACAGACCCAAAGGT 259

QY 121 AC 122
Db 258 AC 257

RESULT 5
AAA89693/c
ID AAA89693 standard; cDNA; 447 BP.
XX
AC AAA89693;
XX
XX 08-JAN-2001 (first entry)
XX
XX Mouse Exo103 nucleotide sequence.
XX
XX Exocytosis pathway protein; Exo protein; antiparkinsonian; antidiabetic;
KW anti-allergic; antiasthmatic; nootropic; neuroprotective; anticonvulsant;
KW vulnary; asthma; inflammation; allergy; Chediak-Higashi syndrome; CHS;
KW Alzheimer's disease; Parkinson's disease; Huntington's disease; diabetes;
XX digestion disorder; wound healing disorder; gene therapy; ss.
XX
OS Mus sp.
XX
XX WO200043419-A2.
XX
XX 27-JUL-2000.

XX PF 20-JAN-2000; 2000WO-US01431.
XX PR 20-JAN-1999; 99US-0116534.
XX PR 26-JAN-1999; 99US-0117274.
XX PR 26-JAN-1999; 99US-0117308.
XX PR 26-JAN-1999; 99US-0117309.
XX PR 26-JAN-1999; 99US-0117312.
XX PR 01-FEB-1999; 99US-0118177.
XX PR 01-FEB-1999; 99US-0118178.
XX PR 01-FEB-1999; 99US-0118179.
XX PR 09-FEB-1999; 99US-0119286.
XX PR 11-FEB-1999; 99US-0119998.
XX PR 11-FEB-1999; 99US-0119759.
XX PA (RIGE-) RIGEL PHARM INC.
XX PI Luo Y;
XX DR WPI; 2000-482908/42.
XX PR New nucleic acids encoding Exo proteins which are useful in the
PT diagnosis, treatment or prevention of exocytosis-mediated disorders
PT such as asthma, inflammation and allergies -
XX
XX PS Disclosure; Page 271; 305pp; English.
XX CC The present sequence encodes a polypeptide which is associated with
CC the exocytosis pathway. cDNA molecules encoding proteins involved in
CC exocytosis have been isolated by yeast one-hybrid and two-hybrid
CC screening. Novel proteins, termed Exo proteins, have been identified that
CC interact with known exocytosis-associated proteins such as GS27, alpha
CC snap, unc18-1, vamp3, snap-23, and the rab family of proteins.
CC Exo proteins and their agonists and antagonists are useful in the
CC diagnosis, treatment or prevention of exocytosis-mediated disorders
CC such as asthma, inflammation, allergies, Chediak-Higashi Syndrome
CC (CHS), Alzheimer's disease, Parkinson's disease, Huntington's disease,
CC diabetes, digestion disorders and wound healing disorders.
CC The nucleic acids, antagonists or agonists of Exo proteins are useful
CC in gene therapy. The nucleic acids are also useful for generating
CC transgenic or knock-out animals which can be used in the
CC development and screening of therapeutically useful reagents.
XX
XX SQ Sequence 447 BP; 102 A; 95 C; 93 G; 157 T; 0 other;
Query Match 78.5%; Score 96.6; DB 21; Length 447;
Best Local Similarity 91.9%; Pred. No. 1.2e-25;
Matches 113; Conservative 0; Mismatches 9; Indels 1; Gaps 1;
QY 1 CAATTGAAAAGTTTGTTC-TAGTGGTGAAGGCCCACTGTGTTCTTCCAGTGA 59
DB 379 CAATTGAAAAGTTTGTTC-TAGTGGTGAAGGCCCACTGTGTTCTTCCAGTGA 320
QY 60 GTTAGGTTGTACAGACGGGTTAGCACTAGCGCTTGAAGAACCTTCACAGACCCCAAGG 119
DB 319 GTTAGGTTGTACAGACGGGTTAGCACTAGCGCTTGAAGAACCTTCACAGACCCCAAGG 260
QY 120 TAC 122
DB 259 AAC 257
RESULT 6
AAH22400/c
ID AAH22400 standard; DNA; 352 BP.
XX
XX AC AAH22400;
XX AC
XX DT 22-AUG-2001 (first entry)
XX DE Human rac1 gene related nucleotide sequence #4.
XX DE Identification; toxic; hepatotoxic; differential gene expression;
KW

KW NSAID; non-steroidal antiinflammatory drug; ds.
XX OS Homo sapiens.
XX PN WO200138579-A2.
XX PD 31-MAY-2001.
XX PF 21-NOV-2000; 2000WO-US32049.
XX PR 22-NOV-1999; 99US-0166923.
XX PR 18-FEB-2000; 2000US-0183531.
XX PR 20-NOV-2000; 2000US-0717321.
XX PA (CURA-) CURAGEN CORP.
XX PI Gould-Rothberg BE, Dipippo VA, Ramseh TM, Gerwein RW;
XX DR WPI; 2001-355948/37.
XX PR Screening hepatotoxic agent comprises contacting test cell population
PT expressing RISKMARKER or INJURYMARKER with agent, comparing expression
PT with reference population and identifying difference in expression
PT levels -
XX
XX PS Disclosure; Page 8-9; 76pp; English.
XX CC The present invention describes a method of screening a test agent for
CC hepatotoxicity. The method comprises: (a) providing a test cell
CC population comprising a cell capable of expressing one or more nucleic
CC acid sequences selected from the group consisting of RISKMARKER 1-8
CC and INJURYMARKER 1-10; (b) contacting the test cell population with a
CC test agent; (c) measuring expression of one or more of the nucleic
CC acid sequences in the test cell population; (d) comparing the
CC expression of the nucleic acid sequence in the test cell population to
CC the expression of the nucleic acid sequence in an reference cell
CC population comprising at least one cell whose exposure status to a
CC hepatotoxic agent is known; and (e) identifying a difference in
CC expression levels of the RISKMARKER or INJURYMARKER sequences, if
CC present, in the test cell population and reference cell population.
CC The method is useful for identifying a hepatotoxic agent. The present
CC sequence is given in the exemplification of the present invention.
XX
XX SQ Sequence 352 BP; 86 A; 77 C; 72 G; 117 T; 0 other;
Query Match 64.7%; Score 79.6; DB 22; Length 352;
Best Local Similarity 83.1%; Pred. No. 2.1e-19;
Matches 103; Conservative 0; Mismatches 19; Indels 2; Gaps 1;
QY 1 CAATTGAAAAGTTTGTTC-TAGTGGTGAAGGCCCACTGTGTTCTTCCAGTGA 60
DB 216 CAATTGAAAAGTTTGTTC-TAGTGGTGAAGGCCCACTGTGTTCTTCCAGTGA 157
QY 61 TTAGGTTGTACAGACGGGTTAGCACTAGCGG--TTGACAGAACCTTCACAGACCCCAAG 118
DB 156 TTAGGTTGTACAGACCATGTCAGCACTAGCAGATTACAGAACCTTCACAGACCCCAAG 97
QY 119 GTAC 122
DB 96 GAAC 93
RESULT 7
ABX83623/c
ID ABX83623 standard; cDNA; 1232 BP.
XX
XX AC ABX83623;
XX AC
XX DT 14-AUG-2002 (first entry)
XX DE Human cDNA differentially expressed in granulocytic cells #194.
XX DE Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW

KW viral infection; parasitic infection; protozoal infection;
KW fungal infection; sterile inflammatory disease; psoriasis;
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW adult respiratory distress syndrome; inflammatory bowel disease;
KW Crohn's disease; ulcerative colitis; periodontal disease;
KW granulocyte activation; chronic inflammation; allergy.
XX Homo sapiens.
XX WO200228999-A2.
XX 11-APR-2002.
XX 03-OCT-2001; 2001WO-US30821.
XX 03-OCT-2000; 2000US-237189P.
XX (GENE-) GENE LOGIC INC.
XX Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX WPI; 2002-435328/46.
XX Detecting granulocyte activation by detecting differential expression
XX of genes associated with granulocyte activation, which serves as
XX diagnostic markers that is useful for monitoring disease states and
XX drug toxicity -
XX Claim 1; SEQ ID NO 194; 114pp; English.
XX The invention relates to detecting (M1) granulocyte (GC) activation
XX (GCA), by detecting the level of expression of gene(s) (Gs) identified by
XX DNA chip analysis as given in the specification, and comparing
XX the expression level to an expression level in an unactivated
XX GC, where differential expression of Gs is indicative of GCA.
XX Also included are modulating (M2) GA by contacting GC with an agent
XX that alters the expression of at least one gene in Gs; (2) screening (M3)
XX for an agent capable of modulating GCA or an inflammation (especially
XX chronic) in a tissue, an allergic response in a subject, exposure of a
XX subject to a pathogen or sterile inflammatory disease using the
XX gene expression profile; (3) detecting (M4) an inflammation (especially
XX chronic) in a tissue, an allergic response in a subject, exposure of a
XX subject to a pathogen or sterile inflammatory disease, by detecting the
XX level of expression in a sample of the tissue of gene(s) from Gs, where
XX the level of expression of the gene is indicative of inflammation;
XX (4) treating (M5) an inflammation (especially chronic) or in a tissue,
XX an allergic response in a subject, exposure of a subject to a pathogen
XX or sterile inflammatory disease, by contacting a tissue having
XX inflammation with an agent that modulates the expression of gene(s)
XX from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
XX modulating GA; M3 is useful for screening an agent capable of modulating
XX GCA preferably in an inflammation in a tissue; M4 is useful for
XX detecting an inflammation (especially chronic) in a tissue, an allergic
XX response in a subject, exposure of a subject to a pathogen or sterile
XX inflammatory disease (e.g. psoriasis, rheumatoid arthritis, renal
XX glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
XX reperfusion injury, ARDS, adult respiratory distress syndrome,
XX inflammatory bowel disease, Crohn's disease, ulcerative colitis,
XX periodontal disease; also bacterial infection, viral infection,
XX parasitic infection, protozoal infection, fungal infection and M5 is
XX useful for treating one of the above conditions. The present
XX sequence represents a gene differentially expressed in granulocytes.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 1232 BP; 329 A; 240 C; 251 G; 411 T; 1 other;
Query Match 64.7%; Score 79.6; DB 24; Length 1232;
Best Local Similarity 83.1%; Pred. No. 3.4e-19;
Matches 103; Conservative 0; Mismatches 19; Indels 2; Gaps 1;

QY 1 CAATTGAAAAAGTTTGTCTAGTGTCAAGGCCAACACATGTCTTTCCAGTGAG 60
DB 1135 CAATTGAAAAATTCGTTGTAGTGTCAAGGGTCCACGCTGTTATCTCGCCAGTGAG 1076
QY 61 TTAGGTTGTACAGACGGGTTAGCACTAGCGC--TTGACAGAACCTCAGACCCCAAG 118
DB 1075 TTAAGTTGTACAGAACATCGTACGCACTAGCAGTTTACAGAACCTCAGACCCCAAG 1016
QY 119 GTAC 122
DB 1015 GAAC 1012
RESULT 8
ABN95134/C
ID ABN95134 standard; DNA; 1232 BP.
XX AC ABN95134;
XX DT 13-AUG-2002 (first entry)
XX DE Gene #1632 used to diagnose liver cancer.
XX KW Gene: liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW metastatic liver tumour; cytostatic; expression profile; disease state;
KW disease progression; drug toxicity; drug efficacy; drug metabolism.
XX OS Homo sapiens.
XX PN WO200229103-A2.
XX PD 11-APR-2002.
XX PF 02-OCT-2001; 2001WO-US30589.
XX PR 02-OCT-2000; 2000US-237054P.
XX (GENE-) GENE LOGIC INC.
XX Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX WPI; 2002-426119/45.
XX Diagnosing and detecting the progression of liver cancer,
XX hepatocellular carcinoma or metastatic liver tumor in a patient,
XX involves detecting the level of expression of two or more genes in a
XX liver tissue sample -
XX Claim 1; SEQ ID NO 1632; 298pp; English.
XX The invention relates to a novel method for diagnosing and detecting the
XX progression of liver cancer, hepatocellular carcinoma or metastatic liver
XX tumor in a patient, and differentiating metastatic liver cancer from
XX hepatocellular carcinoma in a patient, involving detecting the level of
XX expression of two or more genes represented in ABN9503-ABN97455 in a
XX tissue sample. The method of the invention has hepatotropic, and
XX cytostatic activity. The method is useful for diagnosing and detecting
XX the progression of liver cancer, hepatocellular carcinoma and metastatic
XX liver carcinoma in a patient. The method is useful for identifying
XX expression profiles which serve as useful diagnostic markers as well as
XX markers that can be used to monitor disease states, disease progression,
XX drug toxicity, drug efficacy and drug metabolism.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 1232 BP; 329 A; 240 C; 251 G; 411 T; 1 other;
Query Match 64.7%; Score 79.6; DB 24; Length 1232;
Best Local Similarity 83.1%; Pred. No. 3.4e-19;
Matches 103; Conservative 0; Mismatches 19; Indels 2; Gaps 1;

QY 1 CAATTGAAAAGTTTGTCTAGTGTGCTGAAAGGCCCAACACTGTTCTTCCAGTGAG 60
|||||
Db 1135 CAATTTAAAATTTCTGTGTAGTGTGCTGAAAGGTCCTCCACGCTGATTCTCCCAAGTGAG 1076
|||||
QY 61 TTAGTTTGTACAGAACGGCGTTAGCACTAGCGC--TTGACAGAACTCTACAGACCCAAAG 118
|||||
Db 1075 TTAAGTTGTACAGAACATCGTCAGCACTAGCACAGTTTACAGAACTCTACAGACCCAAAG 1016
|||||
QY 119 GTAC 122
|||
Db 1015 GAAC 1012
|||
RESULT 9
AA565569/c
ID AA565569 standard; cDNA; 1318 BP.
XX AC
XX AA565569;
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #1373.
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Drmanac RT, Liu C, Tang YT;
PI
XX
XX WPI; 2001-639362/73.
DR P-PSDB; ABG01382.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 1; SEQ ID No 1373; 103pp; English.
PS
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AA564197-AA594564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

SQ Sequence 1318 BP; 346 A; 268 C; 279 G; 424 T; 1 other;
Query Match 64.7%; Score 79.6; DB 23; Length 1318;
Best Local Similarity 83.1%; Pred. No. 3.5e-19;
Matches 103; Conservative 0; Mismatches 19; Indels 2; Gaps 1;
QY 1 CAATTGAAAAGTTTGTCTAGTGTGCTGAAAGGCCCAACACTGTTCTTCCAGTGAG 60
|||||
Db 1221 CAATTTAAAATTTCTGTGTAGTGTGCTGAAAGGTCCTCCACGCTGATTCTCCCAAGTGAG 1162
|||||
QY 61 TTAGTTTGTACAGAACGGCGTTAGCACTAGCGC--TTGACAGAACTCTACAGACCCAAAG 118
|||||
Db 1161 TTAAGTTGTACAGAACATCGTCAGCACTAGCACAGTTTACAGAACTCTACAGACCCAAAG 1102
|||||
QY 119 GTAC 122
|||
Db 1101 GAAC 1098
|||
RESULT 10
AAF18236/c
ID AAF18236 standard; DNA; 2051 BP.
XX AC
XX AAF18236;
XX
XX 14-MAR-2001 (first entry)
XX
XX Lung cancer associated polynucleotide sequence SEQ ID 255.
DE Human; lung cancer associated protein; neuroprotective; cytostatic;
KW cardioactive; immunomodulatory; muscular active; vulnerary;
KW gastrointestinal; nephrotropic; anti-infective; gynecological;
KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW proliferative disorder; wound healing; infectious disease; ds.
XX
XX Homo sapiens.
OS
XX WO2000055180-A2.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05918.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
XX Ruben SM;
XX
XX WPI; 2000-587514/55.
DR P-PSDB; AAB58360.
XX
XX Lung cancer associated gene sequences, referred to as lung cancer
PT antigens, useful for treatment, prevention, and diagnosis of disorders
PT such as lung cancer -
XX
XX Claim 1; Page 716-717; 1425pp; English.
PS
XX Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC associated proteins and polynucleotide sequences, their agonists, and
CC antagonists may have neuroprotective, cytostatic, cardioactive;
CC immunomodulatory; muscular active general; vulnerary; gastrointestinal
CC general; nephrotropic; anti-infective; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the
CC protein or polynucleotide sequences. The lung cancer associated
CC polynucleotide sequences may be used for detection of lung cancer,
CC chromosome identification, as chromosome markers, and for numerous other
CC diagnostic or research purposes. The proteins may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders. The proteins may also be used in the treatment of wounds and
XX

transcription and addition of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and (2) determining a level or pattern of a nucleic acid in a bovine cell or tissue comprising: (a) incubating a marker nucleic acid (comprising any of the 1512 nucleic acid sequences or its complement or fragment) with a complementary nucleic acid molecule obtained from the bovine cell or tissue, where hybridisation between the marker nucleic acid and the complementary nucleic acid permits the detection of the molecule; and (b) detecting the level or pattern of the complementary nucleic acid, where the detection of the complementary nucleic acid is predictive of the level or pattern of the molecule. The LMPD nucleic acid is used for determining a level or pattern of a molecule in a bovine cell or tissue. It is useful for genome mapping, gene identification and analysis, cattle breeding, preparation of constructs for use in cattle gene expression, or for genetically improving cattle. The present sequence is one of the 1512 bovine LMPD EST (expressed sequence tag) nucleic acids.

Note: The present sequence was not shown in the specification but was obtained in electronic format from the USPTO web site: seqdata.uspto.gov/sequence.html?DocID=20020137139.

Sequence 353 BP; 91 A; 67 C; 85 G; 110 T; 0 other;

Query Match 46.3%; Score 57; DB 25; Length 353;
Best Local Similarity 82.8%; Pred. No. 5.1e-11;
Matches 77; Conservative 0; Mismatches 15; Indels 1; Gaps 1;

QY 1 CAATTGAAAAGTTGTTCTAGTGTGCGAAGGCCCAACACTGTCTTCCAGTGAG 60
|||||
Db 155 CAATTGAAAAGTTGTTCTAGTGTGCGAAGGCCCAACACTGTCTTCCAGTG-G 213
|||||

QY 61 TTAGTTGTACAGAGCGGTTAGCACTAGCGC 93
|||||
Db 214 GTAAGTTGTACAGAACTTCGTTAGCAGGAC 246
|||||

RESULT 13
AAL07296/c
ID AAL07296 standard; DNA; 3740 BP.
XX
AC AAL07296;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 9984.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
XX cancer; gene therapy; ds.
XX Homo sapiens.
XX
XX WO200155320-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01339.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX

PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.

```
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465570/50.
XX
XX Isolated nucleic acid molecule encoding a reproductive system antigen
XX is used in preventing, treating or ameliorating a medical condition -
XX
XX Disclosure; SEQ ID NO 9984; 1297bp + Sequence Listing; English.
XX
XX The present invention provides the protein and coding sequences of a
XX number of human reproductive system related antigens. These can be used
XX in the prevention and treatment of reproductive system disorders,
XX including cancer. The present sequence is a genomic sequence encoding a
XX protein of the invention.
XX
XX Sequence 3740 BP; 1188 A; 752 C; 687 G; 1113 T; 0 other;
SQ
Query Match 31.1%; Score 38.2; DB 22; Length 3740;
Best Local Similarity 78.0%; Pred. No. 0.0012;
Matches 46; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
XX
XX 8 AAAAGTTTGTCTAGTGTGCAAGGCCCAACACTGTGTTCTTGCACGTGAGTTAGGT 66
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX DT 2725 AAAATATTGTCTAGTGTGCAAGGCCCAAGTGTATATCTTGCACGTGCTTAAGGT 2667
```

```
RESULT 14
ABZ04062/c
ID ABZ04062 standard; DNA; 50 BP.
XX
XX AC ABZ04062;
XX
XX DT 09-JAN-2003 (first entry)
XX
XX Human leukocyte gene expression profiling probe SEQ ID NO 4053.
DE
XX
XX T7; leukocyte; gene expression profiling; allograft rejection;
XX atherosclerosis; congestive heart failure; systemic lupus erythematosus;
XX rheumatoid arthritis; osteoarthritis; cytomegalovirus; infection;
XX probe; ss.
XX
XX OS Homo sapiens.
XX
XX FN WO200257414-A2.
XX
XX PD 25-JUL-2002.
XX
XX PF 22-OCT-2001; 2001WO-US47856.
XX
XX PR 20-OCT-2000; 2000US-241994P.
XX
XX PR 08-JUN-2001; 2001US-296764P.
XX
XX PA (BIOC-) BIOCARDIA INC.
XX
XX PI Wohlgemuth J, Fry K, Matcuk G, Altman P, Prentice J, Phillips J;
XX Ly N, Woodward R, Quertermous T, Johnson F;
XX WPI; 2002-636525/68.
XX
XX New system for leukocyte expression profiling, diagnosing a disease, or
XX monitoring (the rate of) progression of a disease, e.g. atherosclerosis
XX or congestive heart failure, comprises diagnostic oligonucleotides -
XX
XX Claim 1; Page 456; 2038pp; English.
XX
XX The invention relates to a system for detecting gene expression, which
XX comprises one or two isolated DNA molecules that detect expression of a
XX gene, where the gene corresponds to any of 8143 oligonucleotides
XX (ABZ00010-ABZ08152) each having 50 base pairs (bp). The system is useful
XX for leukocyte expression profiling. It is particularly useful for
XX diagnosing a disease, monitoring (rate of) progression of a disease,
XX predicting therapeutic outcome, determining prognosis for a patient,
XX predicting disease complications in an individual or monitoring response
XX to treatment in an individual. The diseases include cardiac allograft
XX rejection, kidney allograft rejection, liver allograft rejection,
XX atherosclerosis, congestive heart failure, systemic lupus erythematosus,
XX rheumatoid arthritis, osteoarthritis or cytomegalovirus infection.
XX
XX Sequence 50 BP; 16 A; 15 C; 10 G; 9 T; 0 other;
SQ
Query Match 27.6%; Score 34; DB 24; Length 50;
Best Local Similarity 80.0%; Pred. No. 0.0082;
Matches 40; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
XX
XX 18 TTCTAGTGTGCAAGGCCCAACACTGTGTTCTTGCACGTGAGTTAGGT 67
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX DB 50 TTGTAGTGCTGAGGGTCCACGCTGATTCGCCAGTGAGTTAAGT 1

RESULT 15
AAAL6478
ID AAAL6478 standard; DNA; 605 BP.
XX
XX AC AAAL6478;
XX
XX DT 14-JUN-2000 (first entry)
XX
```

us-09-717-321a-1.rng

Search completed: November 23, 2003, 11:09:27
Job time : 39.0599 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 23, 2003, 09:17:01 ; Search time 360.861 Seconds
(without alignments)
9470.385 Million cell updates/sec

Title: US-09-717-321A-2

Perfect score: 1266

Sequence: 1 ttttttttttttttttcaaa.....gagcaagaataggggatttt 1266

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq 19Jun03.*

1:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1266	100.0	1266	22	AAH22396
2	1017	80.3	1017	22	AAH22397
3	447.6	35.4	2051	21	AAH18236
4	441.8	34.9	28567	25	ABT17030
5	420.2	33.2	1232	24	ABK83623
6	420.2	33.2	1232	24	ABN95134
7	348	27.5	348	22	AAH22399
8	303	23.9	1318	23	AA865569

C	9	290.6	23.0	1022	22	AAH22398	Human rac1 gene re
C	10	230.6	18.2	447	21	AAA89693	Mouse Exol03 nucle
C	11	215.8	17.0	422	21	AAA89694	Mouse rab2 nucleot
C	12	193.2	15.3	605	24	ABL38065	Human colon tumour
C	13	187.8	14.8	352	22	AAH22400	Human rac1 gene re
C	14	187.2	14.8	605	21	AAA16478	Human colon cancer
C	15	172.2	13.6	432	25	ABX42447	Bovine EST associa
C	16	164	13.0	2167	24	ABZ35389	Human gene express
C	17	163.6	12.9	353	25	ABX46267	Bovine EST associa
C	18	126.4	10.0	285	24	ABL71132	Corn tasse1-derive
C	19	123.2	9.7	416	25	ABX43616	Bovine EST associa
C	20	123	9.7	123	22	AAH22395	Human rac1 genomic
C	21	116.4	9.2	3740	22	AAH07296	Human reproductive
C	22	98.2	7.8	581	24	ABV87661	Human colon cancer
C	23	95.4	7.5	201	25	ABX39192	Bovine EST associa
C	24	91.6	7.2	565	24	ABV86985	Human colon cancer
C	25	87.8	6.9	506	22	AAH08004	Human breast cance
C	26	87.6	6.9	470	22	AAH18145	Human breast cance
C	27	85.6	6.8	466	22	AAH25750	Human breast cance
C	28	85.6	6.8	495	24	ABJ37914	Human colon tumour
C	29	68	5.4	588	21	AAA16128	Human colon cancer
C	30	67	5.3	594	24	ABN60324	Human cancer relat
C	31	63.8	5.0	68	16	AAH22394	Human gene signatu
C	32	62.2	4.9	426	25	ABX42177	Bovine EST associa
C	33	61.4	4.8	378	21	AAH06949	Human secreted pro
C	34	60.8	4.8	64	25	ABZ78455	Tumour suppression
C	35	60.8	4.8	64	25	ABZ09002	Human oligonucleot
C	36	60	4.7	404	25	ABX49041	Bovine EST associa
C	37	52.4	4.1	162	24	ABX45998	cDNA encoding colo
C	38	52.2	4.1	424	25	ABX46053	Bovine EST associa
C	39	50	3.9	50	24	ABZ01605	Human leukocyte ge
C	40	49.8	3.9	5852	12	AAQ11710	Dictyostellium plas
C	41	49	3.9	534	23	ABV54457	Human prostate exp
C	42	48.2	3.8	277	25	ABX47508	Bovine EST associa
C	43	48	3.8	640681	24	ABA92787	Buchneca sp. genom
C	44	47.8	3.8	760	22	ABA83286	Human secreted pro
C	45	47.8	3.8	2218	20	AAH84503	Human secreted pro

ALIGNMENTS

RESULT 1
AAH22396
ID AAH22396 standard; DNA; 1266 BP.
XX
XX
AAH22396;
XX
XX
DT 22-AUG-2001 (first entry)
XX
DE Human rac1 contig SEQ ID NO:2.
XX
XX
KW Identification; toxic; hepatotoxic; differential gene expression;
KW NSAID; non-steroidal antiinflammatory drug; ds.
XX
OS Homo sapiens.
XX
PN WO200138579-A2.
XX
PD 31-MAY-2001.
XX
PF 21-NOV-2000; 2000WO-US32049.
XX
PR 22-NOV-1999; 99US-0166923.
PR 18-FEB-2000; 2000US-0183531.
PR 20-NOV-2000; 2000US-0717321.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Gould-Rothberg BE, Dipippo VA, Ramseh TM, Gerwein RW;
XX WPI; 2001-355948/37.


```

XX WPI; 2001-355948/37.
XX
XX Screening hepatotoxic agent comprises contacting test cell population
XX expressing RISKMARKER or INURYMARKER with agent, comparing expression
XX PT with reference population and identifying difference in expression
XX PT levels
XX
XX Disclosure; Page 7-8; 76pp; English.
XX
XX The present invention describes a method of screening a test agent for
XX hepatotoxicity. The method comprises: (a) providing a test agent for
XX population comprising a cell capable of expressing one or more nucleic
XX acid sequences selected from the group consisting of RISKMARKER 1-8
XX and INURYMARKER 1-10; (b) contacting the test cell population with a
XX test agent; (c) measuring expression of one or more of the nucleic
XX acid sequences in the test cell population; (d) comparing the
XX expression of the nucleic acid sequence in the test cell population to
XX the expression of the nucleic acid sequence in an reference cell
XX population comprising at least one cell whose exposure status to a
XX hepatotoxic agent is known; and (e) identifying a difference in
XX expression levels of the RISKMARKER or INURYMARKER sequences, if
XX present, in the test cell population and reference cell population.
XX The method is useful for identifying a hepatotoxic agent. The present
XX sequence is given in the exemplification of the present invention.
XX
XX Sequence 1017 BP; 245 A; 245 C; 216 G; 311 T; 0 other;
XX
Query March 80.3%; Score 1017; DB 22; Length 1017;
Best Local Similarity 100.0%; Pred. No. 1e-244;
Matches 1017; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 245 GAACCTCACAGACCCAGAGGTACCGGAGCATGTGTCGCGTGGTGAGGTCTAGAGGG 304
XX
XX 1017 GAACCTCACAGACCCAGAGGTACCGGAGCATGTGTCGCGTGGTGAGGTCTAGAGGG 958
XX
XX 305 GCGGCATCAATCACATGACAGTGTGGTACTCTGCGAGAGCATGTGTTTCAAGATATC 364
XX
XX 957 GCGGCATCAATCACATGACAGTGTGGTACTCTGCGAGAGCATGTGTTTCAAGATATC 898
XX
XX 365 TAAATAGTTTAAATCTGTAAGCCGAGCAGTGTATTTCTACCCAGTACTAGAAA 424
XX
XX 897 TAAATAGTTTAAATCTGTAAGCCGAGCAGTGTATTTCTACCCAGTACTAGAAA 838
XX
XX 425 ACGAAGGAGCAGTACTGACCTGAGTAAGAGAGGTGAAACAGGAAACGCACTTCTACT 484
XX
XX 837 ACGAAGGAGCAGTACTGACCTGAGTAAGAGAGGTGAAACAGGAAACGCACTTCTACT 778
XX
XX 485 ATCTACCAAAAAATCTCCGAATGATTTATCAGAAAGATCTTATAGTACAGGTGAGACAT 544
XX
XX 777 ATCTACCAAAAAATCTCCGAATGATTTATCAGAAAGATCTTATAGTACAGGTGAGACAT 718
XX
XX 545 ATTGCTCGTTAAGAGGGGTCTTAAAGAAAGCACTTGCTTAAGTTAGCAACTGTGAGGA 604
XX
XX 717 ATTGCTCGTTAAGAGGGGTCTTAAAGAAAGCACTTGCTTAAGTTAGCAACTGTGAGGA 658
XX
XX 605 TGCCGAGTTTAAATATGAGTCTAACGCCCATCTCGGGAGGAGGACAGAGGGGAGGGGG 664
XX
XX 657 TGCCGAGTTTAAATATGAGTCTAACGCCCATCTCGGGAGGAGGACAGAGGGGAGGGGG 598
XX
XX 665 GCTCAAGAGAGACATGATAAGATCGGCCATTTGTCTATCTACTGTTTACAGAAATTAAC 724
XX
XX 597 GCTCAAGAGAGACATGATAAGATCGGCCATTTGTCTATCTACTGTTTACAGAAATTAAC 538
XX
XX 725 CGTTTAAAGAGCTTACCCGTGACACTTTTATTCAGTTGAATTAATCTCAATGTAACAATGAG 784
XX
XX 537 CGTTTAAAGAGCTTACCCGTGACACTTTTATTCAGTTGAATTAATCTCAATGTAACAATGAG 478
XX
XX 785 TGTAATATTAATCTACTTCAATATAGTCAAAATAGTGTCTCTCTTTGATGAGCTG 844
XX
XX 477 TGTAATATTAATCTACTTCAATATAGTCAAAATAGTGTCTCTCTTTGATGAGCTG 418
XX
XX 845 TGTTTACACACTCCACCCAGCACACCCAGCACTAGGAACAGATATCTTGGTAGAGGCA 904
XX

```

RESULT 3

AAFI8236/c

ID AAF18236 standard; DNA; 2051 BP.

XX AC AAF18236;

XX DT 14-MAR-2001 (first entry)

XX DE Lung cancer associated polynucleotide sequence SEQ ID 255.

XX DE Human; lung cancer associated protein; neuroprotective; cytostatic;

XX KW cardiocactive; immunomodulatory; muscular active; vulnery;

XX KW gastrointestinal; nephrotropic; antinefactive; gynecological;

XX KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;

XX KW proliferative disorder; wound healing; infectious disease; ds.

XX OS Homo sapiens.

XX PN WO200055180-A2.

XX PD 21-SEP-2000.

XX PF 08-MAR-2000; 2000WO-US05918.

XX PR 12-MAR-1999; 99US-0124270.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PA (ROSE/) ROSEN C A.

XX PI Ruben SM;

XX DR WPI; 2000-587514/55.

XX DR P-PSDB; AAB58360.

XX PT Lung cancer associated gene sequences, referred to as lung cancer

XX PT antigens, useful for treatment, prevention, and diagnosis of disorders

XX PT such as lung cancer -

XX PS Claim 1; Page 716-717; 1425pp; English.

XX CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer

XX CC associated proteins represented in AAB58106 - AAB58548. Lung cancer

XX CC associated proteins and polynucleotide sequences, their agonists, and

XX CC antagonists may have neuroprotective; cytostatic; cardiocactive;

XX CC immunomodulatory; muscular active general; vulnery; gastrointestinal

CC general; nephrotropic; anti-infective; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the
CC protein or polynucleotide sequences. The lung cancer associated
CC polynucleotide sequences may be used for detection of lung cancer,
CC chromosome identification, as chromosome markers, and for numerous other
CC diagnostic or research purposes. The proteins may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders. The proteins may also be used in the treatment of wounds and
CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC peptide AAF58549 are used in the course of the invention for the
CC identification and characterization of the polynucleotide and protein
CC sequences.
XX
SQ Sequence 2051 BP; 570 A; 430 C; 433 G; 612 T; 6 other;
Query Match 35.4%; Score 447.6; DB 21; Length 2051;
Best Local Similarity 68.2%; Pred. No. 5.4e-102;
Matches 876; Conservative 0; Mismatches 324; Indels 84; Gaps 15;
QY 33 TTTTCTTTTCTTTTATGATCAAGGATTTTAAAGTCAATACATGCAAAACATCTGC 92
Db TTTTCTTTTCTTTTATGATCAAGGATTTTAAAGTCAATACATGCAAAACATATGC 1950
QY 93 TAACTCATTAGCAAAAGATCAATGTAAGAACTCCCAATCTCGCACTGCAATTGA 152
Db TAAATGATTTAGCAAAAGATCAATGTAAGAACTCCCAATCTCGCACTGCAATTGA 1890
QY 153 AAAAAGTTTCTTAGTGGTGAAGGCCCAACACTGTGTCTTGTCCAGTGTAGTT 212
Db AAAAATCTGTTGTAGTGGTGAAGGTCCTCCAGCTGTATCTCGCAGTGTAGTT 1830
QY 213 GTACAGAACGGCTTAGCTAGCTAGCCG--TTGACAGAACTTCAGACCCAAAG-- 264
Db GTACAGAACATCTCAGCACTAGCAGAGTTTACAGAACTTCAGACCCAAAGAAACATC 1770
QY 265 -----TACCGAGACATGTCGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 309
Db AATPAGCAAGGCACTACAGAGGCGGTGTGTCGCGGTGGTGGTGGTGGTGGTGGTGG 1711
QY 310 ATCAATCAATGACAGTGTGTGCTGCTGCAAGACAGTGTG----- 352
Db ATTGGTCAAGTACAGTGTGCTGCTGCTGCAAGACAGTGTGTTAAGAGGTTTCATAGT 1651
QY 353 TTTTCAATATCAATATGATTTTAAACTGTAAAGCGGACGAGTGTCTACACCC 412
Db TTAAGAATTTATCAAAATATTTTAAACTGTAAAGCTGCAACACATGATTTTACACCT 1591
QY 413 AGTTACTAGAAACGAGGAGGACACTAGTACGCTAGTAAAGAGGTCAAAACAGGAA 472
Db AGTTACTAGAAACGAGGAGGACACTAGTACGCTAGTAAAGAGGTCAAAACAGGAA 1536
QY 473 CGCATTCTTACTATCTACCAAAAATCTCCGAATGATTTATCAGAAAGATCTTATAGTA 532
Db AGCATTCTTACTATCTACCAAAAATCTCCGAATGATTTATCAGAAAGATTTTATAGTA 1476
QY 533 CAGCTCAGACATATGCTGCTTAAAGAGGGGCTCTAAAGAAAGCACTGCTTAAGTTAG 592
Db CAAG-GAGGCAATATGCTCAGTCAAGAGGGGCTCTTAAAGAAAGCACTTACTAAGTTAG 1417
QY 593 CAATGTTGAGGATGGCCAGTTTAAATATGACTCAACGCCCATCTCGGAGGAGACAGCA 652
Db CGATTAACAGAACACAGTTTAAAGATGAATTAATGATGATTAATGATGATGATGATGAT 1357
QY 653 GGGGGAAGGGGGCTCAAGAGACACTGATAGATCGGCCATTTGTCATCTACTGTTTG 712
Db GGTGTAAGAGAGAGAGAAAGCTTAAAGAAACATTTCTCTGATTAATACCAACCTTCTTTC 1297
QY 713 ACAGAAATTAACCGTTAAAGAGCTTTACCGGTGACACTTTTATTCAGTTGAA---TTACT 769
Db TCATCTACTGATTTGACAGAAATTAACCTTTTACCGGTGACACTTTTATTCAGTTGAA 1237
QY 770 CCATGTACATGTAGTGTAAATTAATCTCTACTTTTATA-TTAGTCAAAATACTGTCTGTC 828

Db 1236 CTTTCACTAGTGTAGTA---AATCTCACTTGTATTTTGTCAAAATACTGTCTTTG 1181
QY 829 TCCCTTTGATGAGTGTGTTTTCACACTCCACCAGCACCCAGCACTAGGAACAGAA 888
Db TCCCTTTGATGAGTGTGTTTTCACACTCCACCAGCACCCAGCACTAGGAACAGAA 1133
QY 889 TACTTTGTTAGAGGCAACACAGGAGCCAGAGTTCTGTTCAAAAGCCTGCAAGCCGCTCA 948
Db TTTCTTCAATTAGAGGAATA-----GCAGTTCTGTTCAAAATCTCGCAAAAGCTGCTCA 1079
QY 949 GCTGTGATTTTATAGAACTCACTATGAAATCAAGAGCAGAGCTGTTACACCCATCGTGA 1008
Db GAAATCTGCTATGAAATCAAAAGACTGATCCAAAGAGCTGAGTGTCTACGCTCACTCC 1019
QY 1009 CGTACAGTACAAAGTTTACGTAATGAGCATGGGCTGATAGTTTACAGTGGTTTACATGGC 1068
Db ATTACAGTACAAAGTTTATGTCGGGAAACAGCTGCTGCTTAACTCACTGGTGTGATGGC 959
QY 1069 AGCGTGTCAATTAAGAGGCTGTGCTGTACACAGGCTCTGG-----GAGCTACGGGAG 1121
Db AAGCGTTCAATCGGAGGCTGTGCTGTACGATCTGAACTACATAGGAGAGCAAG 899
QY 1122 GGTCTGCACC-CCTGAGCCCCAGAGCTGCAGTCTTCTTAAGGACAAAGTCTCTCAACAGC 1180
Db TGTCTGCACCCTCTAACTGCAAGAGCTACCGTCTTCTCAAAAGACGAAGGTCTTTGCAAG 839
QY 1181 TTATGCTTACGTTGTTTCTCAGACAAAC---GCAACTTAGTTCACAAAGTATTTTGGCAAT 1237
Db TTCAAGTGTGCTGTTTCTCGGCAACAAATGCAAGTGTAGTTCAGAAAGTATTTTGGCAAC 779
QY 1238 TCTTAATCTGAGCAAGATAGGGG 1261
Db TCTTAATCTGAGCAAGATAGGGG 755
RESULT 4
ABT17030/c
ID ABT17030 standard; DNA; 28567 BP.
XX AC ABT17030;
XX AC AC
XX DT 03-APR-2003 (first entry)
XX DE Human MP21 gene Rac1 SEQ ID No 4.
XX KW Cytostatic; p21 pathway modulating agent; cancer; angiogenic; apoptotic;
XX KW cell proliferation disorder; MP21; gene; ds.
XX OS Homo sapiens.
XX PN WO2003006990-A1.
XX PD 23-JAN-2003.
XX PF 10-JUL-2002; 2002WO-US21549.
XX PR 12-JUL-2001; 2001US-305017P.
XX PR 10-OCT-2001; 2001US-328491P.
XX PR 15-FEB-2002; 2002US-357452P.
XX PA (EXEL-) EXELIXIS INC.
XX PI Friedman L, Plowman GD, Belvin M, Li D, Funke RP;
XX WPI; 2003-221779/21.
XX DR P-PSDB; ABU19756.
XX PT Identifying candidate p21 pathway modulator, by contacting an assay
XX PT system having modifiers of p21 polypeptide or gene with a test agent to
XX PT provide a reference activity in system and detecting test agent-biased
XX PT activity -
XX

DR WPI; 2002-435328/46.

XX Detecting granulocyte activation by detecting differential expression

PT of genes associated with granulocyte activation, which serves as

PT diagnostic markers that is useful for monitoring disease states and

PT drug toxicity

XX

PS Claim 1: SEQ ID No 194; 114pp; English.

XX

CC The invention relates to detecting (M1) granulocyte (GC) activation

CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by

CC DNA chip analysis as given in the specification, and comparing

CC the expression level to an expression level in an unactivated

CC GC, where differential expression of Gs is indicative of GCA.

CC Also included are modulating (M2) Gs by contacting GC with an agent

CC that alters the expression of at least one gene in Gs; (2) screening (M3)

CC for an agent capable of modulating GCA or an inflammation (especially

CC chronic) in a tissue, an allergic response in a subject, exposure of a

CC subject to a pathogen or sterile inflammatory disease using the

CC gene expression profile; (3) detecting (M4) an inflammation (especially

CC chronic) in a tissue, an allergic response in a subject, exposure of a

CC subject to a pathogen or sterile inflammatory disease, by detecting the

CC level of expression in a sample of the tissue of gene(s) from Gs, where

CC the level of expression of the gene is indicative of inflammation;

CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,

CC an allergic response in a subject, exposure of a subject to a pathogen

CC or sterile inflammatory disease, by contacting a tissue having

CC inflammation with an agent that modulates the expression of gene(s)

CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for

CC modulating Gs; M3 is useful for screening an agent capable of modulating

CC GCA preferably in an inflammation in a tissue; M4 is useful for

CC detecting an inflammation (especially chronic) in a tissue, an allergic

CC response in a subject, exposure of a subject to a pathogen or sterile

CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,

CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal

CC reperfusion injury, ARDS, adult respiratory distress syndrome,

CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,

CC periodontal disease; also bacterial infection, viral infection,

CC parasitic infection, protozoal infection, fungal infection and M5 is

CC useful for treating one of the above conditions. The present

CC sequence represents a gene differentially expressed in granulocytes.

CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic

CC format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences.

XX

QQ Sequence 1232 BP; 329 A; 240 C; 251 G; 411 T; 1 other;

XX

Query Match 33.2%; Score 420.2; DB 24; Length 1232;

Best Local Similarity 67.5%; Pred. No. 3.3e-95;

Matches 848; Conservative 1; Mismatches 324; Indels 84; Gaps 15;

QY 53 ATTCAAGGATTTATTAAAGTCATACATGCAAAACATCTGCTAATCTAGCAAAAGAT 112

DB 1228 WTCCAAGGATTTATTAAAGTCATACATGCAAAACATATGCTAATTCGATAGCAAAAGAT 1169

QY 113 CAATGTAAACAACTCCCAATCTCAACTGCAATTTGAAAGATTTCTTCTAGTGGT 172

DB 1168 CAATGTAAACAACTCCCAATCTCAACTGCAATTTGAAAGATTTCTTCTAGTGGC 1109

QY 173 CGAAGGCCCAACACTGTGTCTTTCGCGAGTGAGTGTAGTGTGTACAGAACGGCGTTAGCAC 232

DB 1108 TGAAGGGTCCACGGTGTATTCTCGCGAGTGAGTGTAGTGTGTACAGAACATCGTCAGCAC 1049

QY 233 TAGGGC--TTGACAGAACCTCAGACGCCCAAAG--TACCG 269

DB 1048 TAGCACAGTTTACAGAACCTTCACAGACCCCAAGGAACATCAATAGGCAAGCGACTACAG 989

QY 270 GAAGCATGTCTCGCGTGGGTGAGTCTAGAGGGGGGGCGGCATCAATCACATGACAGTGT 329

DB 988 GAGCGTGTCTCGCGTGGCGAGGTAAGA--GGGTGAGTATTGGTCAAGTGACAGTGTCT 930

QY 330 GGTACTCTGGCAAGACAGTGTATG-----TTTCAGAAATATCTAAATAG 372

DB 929 GGTAACTCTGCAAGACAGTGTATTAAAGAGGTTCATAGTTTAAAGAAATTATCTTAAATAT 870

QY 373 TTTTAAAACTGTAAAGCCCGCAGACGTGATTTCTACACCCAGTTTACTAGAAAACGAGGG 432

DB 869 TTTTAAAACTATAAAGCTGCAACATGATTTTACACCTAGTTTACTAGAAAACCTAAGGA 810

QY 433 AAGCACTAGTCAGCTGAGTAAAGAGGTGAAACAGGAGCAGCAGCTTCTACTATCTACCA 492

DB 809 AAGCACTTATAGTCTGAATA-----AAGTAACTGGAAAGCAGCTTTTACTAATCGACA 755

QY 493 AAAAAATCTCCGAATGCATTATCAGAAAGATCTTATAGTACAGGTGAGCATATTCTCTCG 552

DB 754 AAAAAACCTTCTAATGCATTATCAGAAAGATTTATAATACAAG--GAGGCATATTGCTCA 696

QY 553 TTAAGAGGGGGTCTTAAAGAAAAGCAGCTGTAAAGTTAGCACTGTAGGAGTGGCCAGT 612

DB 695 GTCAGAGGGGGTCTTATAAGAAAAGCAGCTTACTAAGTTAGCGACTAACAGAACACAGT 636

QY 613 TTAATATGGACTCAACGCCCATCTGGGAGGAGCAGCAGGGGAGGGGGCTCAAGA 672

DB 635 TTAAGATGAATTAATGCCAATTTGGGAGGCGATGGCAGGTGTAAAGAAAAGAAAG 576

QY 673 GAGACACTGATAAGATCGGCCCATTTGTCTACTCTCTTTTACAGAAAATTAAACCGTTAAAA 732

DB 575 CTTAAGAAAACATTTCTCTGATAATACCAACCTTCTTCTCATCATCTACTGCAATTTGACAG 516

QY 733 AGCTTTACCGTGACACTTTTATTCAAGTTGAA---TTACTCCATGTACATAGTGTAA 789

DB 515 AAATTAACCTTTTAGAGTTTTTACCGGTGACACTTTTCTTCTGTACATGTAGTGTAGTGA 457

QY 790 ATTAACTCTACTTTCATA--TTAGTCAAAATACTGTCTGTCTCTTTGATGACGTCGTGTT 848

DB 456 ---AATCTCCACTTCTGATTTTGTCAAAATACTGTCTTCTGCTTTGATCA----- 409

QY 849 TCACACACTCCACCCAGCAGACCCAGCATAGGAAACAGAAATCTTCTGTAGAGGCAACAC 908

DB 408 -CACACACCCACCCCGGCACACCCACAGCTA--AACAGAAATCTTCTTATTAGAGGAAATAG 352

QY 909 AGGAGCCAGAGTCTCTTTCAAGCCCTGCAAGACCGGTGAGTGTATTTTACAGAACTC 968

DB 351 -----CAGTTCTGTTTCAAAATCTCGCAAAAGCTGTGTCAGAAAACCTCGCTATGAATCA 298

QY 969 ACTATGAATCAAAAGAGCAGAGCTGTATACACCCATCTGACGTACAGTACAAAGTTACGT 1028

DB 297 CAAGACTGATCCAAAGAGCTGAGCTGCTACGCTCACTCCATTACAGTACAATGTTATGT 238

QY 1029 AATGACATGGGCTGATAGTTACAGGTGCTTACAGGTGCTTACAGGTGCTTACATTAAGGAGGCT 1088

DB 237 CGGGAACACGTGCTGCTAATCTACTGTGAGTTCATGTCGCAAGCTTTCATTCGGGAGGCT 178

QY 1089 GTGCTGTGTACACAGGTCTGG-----GAGCTACGGGAGGTCTGCACC--CCTGAGCCC 1140

DB 177 GTTCTGCTTTACGCATCTCAGAACTACATAGGAGAGCAAGTGTCTGCACCTCTCACTGC 118

QY 1141 AGAAGCTGAGTCTTCTTAAAGGACAAAGTCTCTCAACAGCTTGTAGTGTACGTGTTCTCA 1200

DB 117 AGAAGCTACCGTCTTCTCAAGACGAAAGTCTTTGCTAAGTTCAGTCTCGGTGTTCTCG 58

QY 1201 GCACAAAC---GCAACTTAGTTCAAGGATTTTGGCAATTTCTTAATCTGAGCAAGA 1254

DB 57 GCACAAACATGCGATGTAGTTTCAAGAGGATTTTGGCACTCTTAATCTGAAACAAGA 1

RESULT 6

ABN95134/c

ID ABN95134 standard; DNA; 1232 BP.

XX

XX ABN95134;

AC AC

DT 13-AUG-2002 (first entry)

XX DE Gene #1632 used to diagnose liver cancer.

XX	Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic; metastatic liver tumour; cytostatic; expression profile; disease state; disease progression; drug toxicity; drug efficacy; drug metabolism.
XX	Homo sapiens.
XX	WO200229103-A2.
XX	11-APR-2002.
XX	02-OCT-2001; 2001WO-US30589.
XX	02-OCT-2000; 2000US-237054P.
XX	(GENE-) GENE LOGIC INC.
XX	Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX	WPI; 2002-426119/45.
XX	Diagnosing and detecting the progression of liver cancer.
XX	hepatocellular carcinoma or metastatic liver tumor in a patient,
XX	involves detecting the level of expression of two or more genes in a
XX	liver tissue sample -
XX	Claim 1; SEQ ID NO 1632; 298pp; English.
XX	The invention relates to a novel method for diagnosing and detecting the
XX	progression of liver cancer, hepatocellular carcinoma or metastatic liver
XX	tumour in a patient, and differentiating metastatic liver cancer from
XX	hepatocellular carcinoma in a patient, involving detecting the level of
XX	expression of two or more genes represented in ABN93503-ABN97455 in a
XX	tissue sample. The method of the invention has hepatotropic, and
XX	cytostatic activity. The method is useful for diagnosing and detecting
XX	the progression of liver cancer, hepatocellular carcinoma and metastatic
XX	liver carcinoma in a patient. The method is useful for identifying
XX	expression profiles which serve as useful diagnostic markers as well as
XX	markers that can be used to monitor disease states, disease progression,
XX	drug toxicity, drug efficacy and drug metabolism.
XX	CC Note: The sequence data for this patent did not form part of the printed
XX	specification, but was obtained in electronic format directly from WIPO
XX	at ftp.wipo.int/pub/published_pct_sequences.
XX	Sequence 1232 BP; 329 A; 240 C; 251 G; 411 T; 1 other;
XX	Query Match 33.2%; Score 420.2; DB 24; Length 1232;
XX	Best Local Similarity 67.5%; Pred. No. 3.3e-95;
XX	Matches 848; Conservative 1; Mismatches 324; Indels 84; Gaps 15;
QY	53 ATTCAAGGATTTATTAAAGTCATACATGCAAAACATCTGCTAACTGCATTAGCAAAAGAT 112
Db	: WTCCAAGGATTTATTAAAGTCATACATGCAAAACATATGCTAATTGCAATTAGCAAAAGAT 1169
QY	113 CAATGTAAAAACACTCCAAATTCGCACTGTCAATGTGAAAAAGTTTCTTAGTGGT 172
Db	1168 CAATGTAAAAACACTCCAAATTCGCACTGTCAATTTAAAAAATCTCTTGTAGTGGC 1109
QY	173 CGAAAGGCCCAACACTGTGTTCTTTCGCAGTGAGTTGTACAGAACGGCGTTAGCAC 232
Db	1108 TGAAGGGTCCACGCTGTATTCTTCGCAGTGAGTTGTACAGAACATCGTCAGCAC 1049
QY	233 TAGCGC--TTGACAGAACCTTCACAGACCCAAAGG-----TACCG 269
Db	1048 TAGCACAGTTTACAGAACCTTCACAGACCCAAAGAACATCAATAGGCAAGCACTACAG 989
QY	270 GAAGCATGTGTCGCGTGGGTGAGGTTAGAGGGGGCGGCATCAATCACATGACAGTGT 329
Db	988 GAGGCGGTGTTCGCGCGTGGGCGAGGTAAGA-GGGTCAGTATTGGTCAAGTCACAGTGC 930
QY	330 GGTACTCTGSCAAGACAGTGATG-----TTTCAGATATCTAAATAG 372
Db	929 GGTAATCTGCAAGACAGTGATGTTAAGAGGTTTCATAGTATTGAAGTAATCTCAAAATAT 870

Db 1254 CAATGTAARAACACTCCACRAATCTGCAACTGTCAAATTTAAATAAATCTGTTGTAGTGGC 1195
QY 173 CGAAGGCCCAACACTGTGTTCTTGGCAGTGTAGTTGTTAAGAACGGCGTTAGCAC 232
Db 1194 TGAAGGGTCCACGCTGTATTTCTGCGCAGTGTAGTTGTTAAGAACCATCGTCAGCAC 1135
QY 233 TAGCGC--TTGACAGAACCTCACAGACCCCAAGG-----TACCG 269
Db 1134 TAGCACAGTTTACAGAACTCACAGACCCCAAGGAAATCAATAGGCAAAAGCACTACAG 1075
QY 270 GAACATGTGTCCGCGTGGTGTAGGTTCTAGAGGGGGCGGCATCAATCACATGACAGTGT 329
Db 1074 GAGCGGTGTGTCGCGTGGCGAGGTAAGA-GGGTCAGTATTGTTCAAGTGACAGTGC 1016
QY 330 GGTACTCTGGACAGACAGTGTATTTCA-----GAATATCTAAATAG 372
Db 1015 GGTAAATCTGGACAGACAGTGTATTTAAGAGGTTTCATAGTTTAAAGAAATATCTAAATAT 956
QY 373 TTTAAAAAAGTAAAGCGGACGACGTGATTTCTACACCCAGTTACTAGRAAAGCAAGGG 432
Db 955 TTTAAAAAATATAAGCTGCACACATGATTTTACCTAGTTACTAGAAATCTAAGGA 896
QY 433 AAGCACTAGTGTAGTGTAGTAAAGGAGTGAAGAACGACGCTTCTACTATCTACCA 492
Db 895 AAGCACTATTAGTCTGTAATAA-----GTAAACATGGAAGCACTTTTACTATCGACA 841
QY 493 AAAAAATCTCGAATGATATATCAGAAGATCTTATAGTACAGTGTACACATATTGTCG 552
Db 840 AAAAAACCTTCTAATGCAITATCAGAAGATTTTATAACAAG-GAGGCATATTGTCTCA 782
QY 553 TTAAGAGGGGCTCTTAAGAAA-----AGCACTTGTAGTTAGCACTGTGAGGATGG 607
Db 781 GTCAAGAGGGGTTCTTTTAGGAAAGGACCTTTACTAAGTTAAGCCACTTAACAGAACAA 722
QY 608 CCAAGTTTAAATATGACTCAA---CGCCCCACTCTGGGAGGAGCAGCAGGGGGAAGGGG 664
Db 721 CCAAGTTTAAAGCATGAATTAATGCCCCAAATTTGGGGGAGGCAATGGCAGGTTGAAGAGA 662
QY 665 GCTCAAGAGA-----GACACTGATAAGATCGGCCANTTGTAT 702
Db 661 AAGGAAAAGGCTTAAGAAAAATTTCTGTGATAATACCAAACTTCTTTCAATCATCTA 602
QY 703 CTACTGTTTACAGAAATTAACCGTTTAAAGCTTTACCGGTGACACTTTTATTCAGTTG 762
Db 601 CTGCAITGGACAGAAATTAACCTTTTAGAGTTTTTACCCCGTGACCACTTTTCATTC 542
QY 763 AATTACTCCATGTACAAATGATGTAAATTAATCTCTACTTTCATATTAGTCMAAATACTG 822
Db 541 CTGTGTACCAATGGTAGTGGTAATCTCCCACTCCGTATTTGTGTTCCA--AAATTACTG 485
QY 823 TCTGTCTCCTTTGATGACGTGTTGTTTCAACACTCCACCAGACACCCACGACTAGGA 882
Db 484 TCTTTGTCTTTGA-----TCCACCACCAACACCCGCCCGGACACCCCCACAGCTAAA 430
QY 883 ACAGAAATCTGTTTAGAGGCAACACAGGAGCCAGAGTCTGTTTAAAGCCTGCAGAGC 942
Db 429 CCAGAAATCTCTATTAGAGAAATACCCAGTTCTGTTTAAATCTCCCGCAAAAGCTGG 370
QY 943 CGGTACGTGGTATTTTAGAGAACTCACTATGAAATCAAGAGCAGAGCTGTTTACACCCA 1002
Db 369 TCAGAAACTCGCTATGAATCAAAAGACTGATCCAAAGAGCTGAGGCTGTACGTTCA 310
QY 1003 TCGTGACGTACAGTACAAAGTTAGTAAAGCATGGGCTG-ATAAGTTACAGTGGCTT 1061
Db 309 --CTCCATTACAGTACAAATTTATGTGCGGAACACGTCGCTGCTAACTCACTGTGAT 252
QY 1062 ACATGGCAGGTGTCAATTAAGAGGCTGTGTTGTACACGGTCTCGGA-----GC 1113
Db 251 CAATGGCAACGCTTCAATTCGGGAGGCTGTTCTGCTTTACGCATCTGAGAACTACATAGGA 192
QY 1114 TAGGGAGGGTCTGCACC-CCTGAGCCCAAGAGTGTGAGTCTTCTTAA--GGACAAAGTC 1170
Db 191 GAGCAAGTGTCTGCACCTCCTAACTGCAGAGAGTACCGTCTTCTCAATGACGAGGTC 132

QY 1171 TCTCAACAGCTTAGTGT--TACGTGTTCTCAGCAACGCAACTTAGTTCAAGGTAT 1228
Db 131 TTTGCAAGTTTCAGTGTGCGTTCGCGCACCAATGCAGTGTAGTTCAAGGTAT 72
QY 1229 TTTGGCAATCTTAACTGTGACCAAGATAGGG 1261
Db 71 TTTGGCACTCTTAACTGTGACCAAGATAGGG 39
RESULT 9
AAH22398/C
ID AAH22398 standard; DNA; 1022 BP.
XX AAH22398;
XX AC
XX AAH22398;
XX 22-AUG-2001 (first entry)
XX Human rac1 gene related nucleotide sequence #2.
DE
XX
XX Identification; toxic; hepatotoxic; differential gene expression;
KW NSAID; non-steroidal antiinflammatory drug; ds.
XX
XX Homo sapiens.
OS
XX
XX WO200138579-A2.
XX
XX 31-MAY-2001.
XX
XX 21-NOV-2000; 2000WO-US32049.
XX
XX 22-NOV-1999; 99US-0166923.
PR 18-FEB-2000; 2000US-018331.
PR 20-NOV-2000; 2000US-0717321.
XX
XX
XX (CURA-) CURAGEN CORP.
XX
XX Gould-Rothberg BE, Dipippo VA, Ramseh TM, Gerwein RW;
XX WPI; 2001-355948/37.
XX
XX Screening hepatotoxic agent comprises contacting test cell population
XX expressing RISKMARKER or INJURYMARKER with agent, comparing expression
XX with reference population and identifying difference in expression
XX levels
XX
XX Disclosure; Page 7-8; 76pp; English.
XX
XX The present invention describes a method of screening a test agent for
XX hepatotoxicity. The method comprises: (a) providing a test cell
XX population comprising a cell capable of expressing one or more nucleic
XX acid sequences selected from the group consisting of RISKMARKER 1-8
XX and INJURYMARKER 1-10; (b) contacting the test cell population with a
XX test agent; (c) measuring expression of one or more of the nucleic
XX acid sequences in the test cell population; (d) comparing the
XX expression of the nucleic acid sequence in the test cell population to
XX the expression of the nucleic acid sequence in an reference cell
XX population comprising at least one cell whose exposure status to a
XX hepatotoxic agent is known; and (e) identifying a difference in
XX expression levels of the RISKMARKER or INJURYMARKER sequences, if
XX present, in the test cell population and reference cell population.
XX The method is useful for identifying a hepatotoxic agent. The present
XX sequence is given in the exemplification of the present invention.
XX
SQ Sequence 1022 BP; 273 A; 208 C; 205 G; 336 T; 0 other;
Query Match 23.0%; Score 290.6; DB 22; Length 1022;
Best Local Similarity 64.5%; Pred. No. 8.8e-63;
Matches 664; Conservative 0; Mismatches 304; Indels 61; Gaps 13;
QY 265 TACCGNAGCATGTCCGGTGGTGGTGTAGAGGGCGGCATCATCATGACA 324
Db 1000 TACAGGAGCGGTGTGTCGGTGGGAGTAAGA-GGGTCAGTATTGTCAGGTGACA 942

QY 325 GTGTTGGTACTCTGGCAGACAGTGTATTTTCAGAA-----TATCTAA 367
 AC |||||
 DT 941 GTGTCGGTAAATCTGGCAGACAGTGTATTTTCAGAAAGGTTCAATGTTTAAGAATTAATCTAA 882
 AC |||||
 QY 368 AATAGTTTAAAACTGTAAGCGCGACAGCTGATTTCTACACCCAGTTACTAGAAAACG 427
 AC |||||
 Db 881 AATATTTTAAAACTGTAAGCGCGACACATGATTTTACACCTAGTTACTAGAAAAC 822
 AC |||||
 QY 428 AAGGGAAGCACTAGTCTAGTGAAGGAGGTGAAACAGGAGCGCTTCTACTATC 487
 AC |||||
 Db 821 AAGGGAAGCACTATTAGTCTCTGAATA-----AAGTAACATGGAAGCACTTTTACTAAT 767
 AC |||||
 QY 488 TACCAAAAAAATCTCCGAATGCAATTATCAGAAAGATCTTATAGTACAGGTCAGACATATT 547
 AC |||||
 Db 766 CGCAAAAAAATCTTCTAATGCAATTATCAGAAAGATTTTATAATACAAG-GAGGCATATT 708
 AC |||||
 QY 548 GTCGTGTAAGAGGGGGTCTTAAAGAAAAGCACTTCTGTAAGTTAGCAACTGTGAGGATGG 607
 AC |||||
 Db 707 GCTCAGTCAGAAGGGGGTCTTATAAGAAAAGCACTTACTAAGTTAGCGACTAACAGAACAA 648
 AC |||||
 QY 608 CAGTTTAAATATGATCAAGCCCACTCTGGGAGGAGACAGCGGGGAGGGGGCT 667
 AC |||||
 Db 647 CAGTTTAAAGATGAATTAATGCCCAATTTGGGGAGGATGCCAGGTGTAAAGAGAAAG 588
 AC |||||
 QY 668 CAAGAGAGACACTGATAAGATCGGCCAATTTGTCTACTCTGTTTGACAGAAATTAACCGT 727
 AC |||||
 Db 587 AAGAGCTTAAAGAAACATTTCTGTATTAATACCAACCTTTCTTTCATCACTACTGCTATT 528
 AC |||||
 QY 728 TAAAGAGCTTACCCGTGACACTTTTATTCAGTTGAATT---ACTCCATGTAACATGTAG 784
 AC |||||
 Db 527 GACAGAAATTAACCTTTTAGAGTTTTTACCCGTGACACTTTCAATTCCTTGTAACAATGTAG 468
 AC |||||
 QY 785 TGTAAATTAATCTCTACTTCATA-TTAGTCAAAATACCTCTCTCTCTTTGATGACGTC 843
 AC |||||
 Db 467 TGTA-----AATCTCACTTCGTATTTTGTCAAATATCTGTCTTTGTCCTTTGATCA--- 416
 AC |||||
 QY 844 GTGTTTACACACTCCACCCAGACACACCCACGACTAGGAACAGAAATCTTCGTTAGAGGC 903
 AC |||||
 Db 415 -----CACACACCCACCCGACACACCCACAGCTA--AACAGAAATCTTTCATTAGAGGA 364
 AC |||||
 QY 904 AACACAGAGCCAGAGTTCTGTTTCAAGCGTCGAGAGCGCGTCACTGCTGATTTTAGAG 963
 AC |||||
 Db 363 AATA-----GCAGTTCTGTCTCAAAATCTCCGCAAAAGCTGGTTCAGAAAACCTCGCTATGA 310
 AC |||||
 QY 964 AACTCACTATGAATCAAGAGCAGAGCTGTTACACCCATCGTACAGTACAGTACAAAT 1023
 AC |||||
 Db 309 AATCAAAAGACTGATCCAAAGAGCTGAGCTGCTACGCTCACTTCCATTAAGTACAGTAAATGT 250
 AC |||||
 QY 1024 TAGCTAATGAGCATGGGCTGATAAGTTACAGGTGCGTTTACATGGCAGCGTGTCAATTAAG 1083
 AC |||||
 Db 249 TATGTCGGGAACACAGTGTCTTAACCTCACTGTTGAGTTCAATGTCGAACGCTTCAATTCGG 190
 AC |||||
 QY 1084 AGGCTGTGCTGTGTACAGCTCTGG-----GAGCTAGGGAGGCTGTGACC-CTTG 1135
 AC |||||
 Db 189 AGGCTGTCTGTGTTTACGATCTGTGAAGTACATAGGAGAGCAAGTGTCTGCACTTCCTTA 130
 AC |||||
 QY 1136 AGCCAGAGAGCTGCACTTCTTAAAGCAAACTCTCAACAGCTTAGTGTACGTGT 1195
 AC |||||
 Db 129 ACTGCAGAGTACCGCTCTTCTCAAGAGCAAGGCTTTTGCAGAGTTCAGTCTCGGTGT 70
 AC |||||
 QY 1196 TCTAGCACAAC---GCAACTTAGTTTCAAGAGTATTTTGGCAATTTCTTAATCTGAGCAA 1252
 AC |||||
 Db 69 TCTCGGCACAACAATGCAAGTGTAGTTTCAAGAGTATTTTGGCAACTCTTAATCTGAACAA 10
 AC |||||
 QY 1253 GAATAGGG 1261
 AC |||||
 Db 9 GAATAGGG 1
 AC |||||

XX AAA89693;
 AC
 XX 08-JAN-2001 (first entry)
 DT
 XX Mouse Exol03 nucleotide sequence.
 DE
 XX
 KW Exocytosis pathway protein; Exo protein; antiparkinsonian; antidiabetic;
 KW antiallergic; antiasthmatic; nootropic; neuroprotective; anticonvulsant;
 KW vulnery; asthma; inflammation; allergy; Chediak-Higashi syndrome; CHS;
 KW Alzheimer's disease; Parkinson's disease; Huntington's disease; diabetes;
 KW digestion disorder; wound healing disorder; gene therapy; ss.
 XX
 OS Mus sp.
 XX WO2000043419-A2.
 XX
 XX 27-JUL-2000.
 PD
 XX
 XX 20-JAN-2000; 2000WO-US01431.
 PF
 XX
 XX 20-JAN-1999; 99US-0116534.
 PR
 XX 26-JAN-1999; 99US-0117274.
 PR
 XX 26-JAN-1999; 99US-0117308.
 PR
 XX 26-JAN-1999; 99US-0117309.
 PR
 XX 26-JAN-1999; 99US-0117312.
 PR
 XX 01-FEB-1999; 99US-0118177.
 PR
 XX 01-FEB-1999; 99US-0118178.
 PR
 XX 01-FEB-1999; 99US-0118179.
 PR
 XX 09-FEB-1999; 99US-0119286.
 PR
 XX 11-FEB-1999; 99US-0119998.
 PR
 XX 11-FEB-1999; 99US-0119759.
 PR
 XX (RIGB-) RIGEL PHARM INC.
 PA
 XX
 XX Luo Y;
 PI
 XX
 XX WPI; 2000-482908/42.
 DR
 XX
 XX
 XX New nucleic acids encoding Exo proteins which are useful in the
 PT diagnosis, treatment or prevention of exocytosis-mediated disorders
 PT such as asthma, inflammation and allergies -
 XX
 XX Disclosure; Page 271; 305pp; English.
 PS
 XX
 XX The present sequence encodes a polypeptide which is associated with
 CC the exocytosis pathway. cDNA molecules encoding proteins involved in
 CC exocytosis have been isolated by yeast one-hybrid and two-hybrid
 CC screening. Novel proteins, termed Exo proteins, have been identified that
 CC interact with known exocytosis-associated proteins such as GS27, alpha
 CC snap, unc18-1, vamp3, snap-23, and the rab family of proteins.
 CC Exo proteins and their agonists and antagonists are useful in the
 CC diagnosis, treatment or prevention of exocytosis-mediated disorders
 CC such as asthma, inflammation, allergies, Chediak-Higashi Syndrome
 CC (CHS), Alzheimer's disease, Parkinson's disease, Huntington's disease,
 CC diabetes, digestion disorders and wound healing disorders.
 CC The nucleic acids, antagonists or agonists of Exo proteins are useful
 CC in gene therapy. The nucleic acids are also useful for generating
 CC transgenic or knock-out animals which can be used in the
 CC development and screening of therapeutically useful reagents.
 CC
 XX Sequence 447 BP; 102 A; 95 C; 93 G; 157 T; 0 other;
 SQ
 Query Match 18.2%; Score 230.6; DB 21; Length 447;
 Best Local Similarity 78.4%; Pred. No. 7e-48;
 Matches 349; Conservative 0; Mismatches 69; Indels 27; Gaps 5;
 QY 77 ATGCAAAACATCTGCTAACTGCTAATGCAAAAGATCAATGTAAAAACACTCCAAATTC 136
 Db 447 ATGCAAAACCTACTGCTAACTGCTAATGCAAAAGATCAATGT-AAAACACTCCAAATTC 389
 QY 137 TGCACACTGCAATTGAAAAAGTTTCTTC-TAGTGTGCAAGGCCCACTGTGTCT 195
 |||||

XX 09-JUN-2000; 2000US-210899P.
PR 20-FEB-2001; 2001US-270216P.
XX (CORI-) CORIXA CORP.
XX Jiang Y, Harlocker SL, Secrist H;
XX WPI; 2002-114514/15.
XX Novel isolated colon tumor polynucleotide differentially expressed in
PT colon tumor or colon metastatic tumor and polypeptides encoded by them,
PT useful for inhibiting development of cancer in patient -
XX
PS Claim 1: SEQ ID 1654; 105pp; English.
XX
XX ABL36412 to ABL38645 represent human colon tumour antigen cDNA clones (I)
CC which were isolated from human colon tumour and colon metastatic tumour
CC cDNA libraries. (I) have cytostatic activity and can be used in vaccine
CC production. (I) can be used for stimulating and/or expanding T cells
CC specific for a tumour protein on contact with the T cells. They are also
CC useful for inhibiting the development of cancer in a patient. (I) can be
CC used as probes or primers for nucleic acid hybridisation, for preparing
CC mutant species primers, or primers for use in genetic constructions. (I)
CC can be used in the diagnosis of a colon tumour.
XX
SQ Sequence 605 BP; 219 A; 109 C; 117 G; 156 T; 4 other;
Query Match 15.3%; Score 193.2; DB 24; Length 605;
Best Local Similarity 70.5%; Pred. No. 1.8e-38;
Matches 371; Conservative 0; Mismatches 115; Indels 40; Gaps 7;
QY 265 TACCGAAGCATGTGTCGGGTGGGTGAGTGTAGAGGGGGCGGATCATCATGACA 324
Db 74 TACAGGAGCGGTGTGTCGGGTGGGCGAGGTAAAGA-GGTCAGTATGGTCAAGTGCCA 132
QY 325 GTGTTGGTACTCTGGCAAGACAGTGTATGTTTCAGAA-----TATCTAA 367
Db 133 GTGTCGGTAACTGCGCAAGACAGTGTATGTTAAAGAGTTTCATAGTTTAAAGATTACTAA 192
QY 368 AATAGTTTAAACTTAAAGCGCAGCAGTGTATGTTTACACCCAGTGTACTAGAAAACG 427
Db 193 AATATTTTAAATACTAAAGCTGCAACACATGATTTTACACCTAGTTACTAGAAAAC 252
QY 428 AAGGAGACACTAGTACAGTGTATGAGGAGGTGAACAGGACGCTTCTACTATC 487
Db 253 AAGGAAGCACTTATAGTCTGAATA-----AAGTAACATGGAAGCACTTTTACTAAT 307
QY 488 TACCAAAAAATCTCCGATGCAATTATCAGAAAGATCTTATAGTACAGGTCAGACATATT 547
Db 308 CGACAAAANAACCTTCTAATGCATTATCAGAAAGATTATATACAAG-GAGGCATATT 366
QY 548 GTCGTTAAGAGGGGGTCTTAAAGAAAGCATCTGCTAAGTTAGCAACTGTGAGGATGG 607
Db 367 GCTCAGTCAGAGGGGTTCTATAGAAAGACACTTACTAAGTTAGGCACTTAACAGAACAA 426
QY 608 CCAAGTTTAAATATGAGTCAACGCCCATCTGGGAGGGACAGCAGG-----GGG 657
Db 427 CCAAGTTTAAAGATGAATTAATGCCAATTTGGGAGGAGCATGGCAGGTGAAGAAAGG 486
QY 658 AAGGGGGGCTCAAGAGAGACACTGATAAGATCGCG-----CATTTGTCACTACTG--TTT 711
Db 487 AAAAGCTTAAAGAAACATTTCTTGATTAATACCAACCTTTCTTTCATCATCTACTGCATT 546
QY 712 GACAGAAATTAACGGTTTAAAGATTTTACCGGTGACACTTTTATTC 757
Db 547 GACAGAAATTAACCTTTTAGAGNTTTTACCCNTGACACTTTTCATTC 592

RESULT 13
AAH22400/c
ID AAH22400 standard; DNA; 352 BP.
XX

AC AAH22400;
XX 22-AUG-2001 (first entry)
XX Human rac1 gene related nucleotide sequence #4.
XX Identification; toxic; hepatotoxic; differential gene expression;
KW NSAID; non-steroidal antiinflammatory drug; ds.
XX Homo sapiens.
XX WO200138579-A2.
XX 31-MAY-2001.
XX 21-NOV-2000; 2000WO-US32049.
XX 22-NOV-1993; 99US-0166923.
PR 18-FEB-2000; 2000US-0183531.
PR 20-NOV-2000; 2000US-0717321.
XX (CURA-) CURAGEN CORP.
XX Gould-Rothberg BE, Dipippo VA, Ramsen TM, Gerwein RW;
XX WPI; 2001-355948/37.
XX Screening hepatotoxic agent comprises contacting test cell population
PT expressing RISKMARKER or INJURYMARKER with agent, comparing expression
PT with reference population and identifying difference in expression
PT levels -
XX Disclosure; Page 8-9; 76pp; English.
PS The present invention describes a method of screening a test agent for
XX hepatotoxicity. The method comprises: (a) providing a test cell
XX population comprising a cell capable of expressing one or more nucleic
CC acid sequences selected from the group consisting of RISKMARKER 1-8
CC and INJURYMARKER 1-10; (b) contacting the test cell population with a
CC test agent; (c) measuring expression of one or more of the nucleic
CC acid sequences in the test cell population; (d) comparing the
CC expression of the nucleic acid sequence in the test cell population to
CC the expression of the nucleic acid sequence in a reference cell -
CC population comprising at least one cell whose exposure status to a
CC hepatotoxic agent is known; and (e) identifying a difference in
CC expression levels of the RISKMARKER or INJURYMARKER sequences, if
CC present, in the test cell population and reference cell population.
CC The method is useful for identifying a hepatotoxic agent. The present
CC sequence is given in the exemplification of the present invention.
XX
SQ Sequence 352 BP; 86 A; 77 C; 72 G; 117 T; 0 other;
Query Match 14.8%; Score 187.8; DB 22; Length 352;
Best Local Similarity 77.8%; Pred. No. 3.4e-37;
Matches 274; Conservative 0; Mismatches 57; Indels 22; Gaps 3;
QY 10 TTTTTCAGTTCACAAAGACATTTTTCATTTTTCATGATTCAGGATTTATTAA 69
Db 352 TTGTTTCATCAAGTCTCAAAAACACCAGTATTACCAGTCGTATGATTCAAGGATTTATTAA 293
QY 70 GTCATCATGCAAAAACATCTGCTAATCTAGTATGCAAAAGATCAATGTAACACATCC 129
Db 292 GTCATCATGCAAAAACATATGCTAATTCATTTAGCAAAAGATCAATGTAACACATCC 233
QY 130 ACATTCCTGCAACTGTCAATTTGAAAAAGTTTCTTAGTGGTCGAAAGGCCCAACACTG 189
Db 232 ACATTCCTGCAACTGTCAATTTAAAAAATCTGTTGTAGTGGCTGAAGGGTCCACGCTG 173
QY 190 TGTTCCTTGCAGTGAAGTTAGGTTGTACAGAACCGGTTAGCACTAGCGC--TTGACAGAA 247
Db 172 TATTCGCGCAGTGAGTTAAGTTGTACAGAACATCGTCAGCACTAGCACAGTTTACAGAA 113
QY 248 CTTACAGACCCCAAGGTAC-----CGGAAGCATGTGTCGCGGTGG 288

Mon Nov 24 15:20:15 2003

CC second nucleic acid molecule comprising any of 15112 nucleotide
CC sequences, appearing as ABX3486-ABX4947, or complements of them.
CC Also included are; (1) a transformed cell having a nucleic acid
CC comprising an LMPD nucleic acid linked to a promoter and a 3' non-
CC translated sequence that functions in the cell to cause termination of
CC transcription and addition of polyadenylated ribonucleotides to a 3' end
CC of the mRNA molecule; and (2) determining a level or pattern of a
CC molecule in a bovine cell or tissue comprising: (a) incubating a marker
CC nucleic acid (comprising any of the 15112 nucleic acid sequences or its
CC complement or fragment) with a complementary nucleic acid molecule
CC obtained from the bovine cell or tissue, where hybridisation between the
CC marker nucleic acid and the complementary nucleic acid permits the
CC detection of the molecule; and (b) detecting the level or pattern of the
CC complementary nucleic acid, where the detection of the complementary
CC nucleic acid is predictive of the level or pattern of the molecule.
CC The LMPD nucleic acid is used for determining a level or pattern
CC of a molecule in a bovine cell or tissue. It is useful for genome
CC mapping, gene identification and analysis, cattle breeding, preparation
CC of constructs for use in cattle gene expression, or for genetically
CC improving cattle. The present sequence is one of the 15112 bovine
CC LMPD EST (expressed sequence tag) nucleic acids.
CC Note: The present sequence was not shown in the specification but
CC was obtained in electronic format from the USPTO web site:
CC seqdata.uspto.gov/sequence.html?DocID=20020137139.
XX

SQ Sequence 432 BP; 123 A; 76 C; 78 G; 155 T; 0 other;

Query Match 13.6%; Score 172.2; DB 25; Length 432;
Best Local Similarity 72.2%; Pred. No. 2.9e-33;
Matches 294; Conservative 0; Mismatches 103; Indels 10; Gaps 5;
QY 379 ACTGTAAAGCGCAGCAGCTGATTTCTA-CACCCAGTTACTAGAAACGAGGGGAAGCA 437
DB 432 AACCATAAAGCTGCAACAGCTGATTTTATACCTAGTTACTAGAAACTAAGGAAGCA 373
QY 438 CTAGTCAGC--TGAGTAAAGGAGGTGAAACAGGAACGACCTTCTACTATCTACCAAAA 495
DB 372 CTTATTAGCTTTGATTAAGCAACATGAAACAGAGGTGCACCTTTACTAACCTACAAAA 313
QY 496 AAATCTCCGAATGCAATTATCAGAAAGATCTTATAGTACAGGTGAGCAGATATTCCTGTTA 555
DB 312 AAATTTTCTAATGCATATTCAGAAAGATTTTATAATACAAG-GAGGCATATTCCTCATTA 254
QY 556 AGAAGGGGTCTTAAGAAAGACACTTGCTAGCTAGCACTGTGAGGATGGCCAGTTTA 615
DB 253 AGAATGAGTTCTATAAGAAAGCACTTACTAGTTAGCAACTATGAGGATGACCAAGTCA 194
QY 616 AATATGGACTCAACGCC---CCATCTGGGGAGGACAGCAGGGGGAAGGGGGCTCAAG 671
DB 193 GAGATGATTAAATGCCCAATTTCAGAAAGGGGTGGCGAGTTTAAGGAAAGCTTAAGAAA 134
QY 672 AGAGACACTGATAGATCGGCCAATTTGTCTACTTG--TTTGACAGAAATTAACCGTTA 729
DB 133 ACACCTACTGATAATACCGACCTTCTTCATCTACCGCAATTTGACAGAAATTAACCTTTT 74
QY 730 AAAGCTTTTACCGTGACACTTTTATTCAGTTGATTTACTTCCATGTA 776
DB 73 AAAAATTTTACCGGTGATGCTTTTATATAGTTTAAAGCTTTACGTGTA 27

Search completed: November 23, 2003, 11:09:34
Job time : 367.861 secs


```

gene expression in liver
Patent: WO 018579-A 2 31-MAY-2001;
Curagen Corporation (US)
FEATURES
Location/Qualifiers
1..1266
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
BASE COUNT
385 a 258 c 285 g 338 t
ORIGIN

```

Query Match	100.0%	Score 1266	DB 5	Length 1266
Best Local Similarity	100.0%	Prod. No. 0		
Matches 1266	Conservative 0	Mismatches 0	Indels 0	Gaps 0
Qy	1	TTTTTTTTTTTTTTTTTTCAAGTCCAAAGACATTTTTTTTTTTTTTTTTTTTATGATCAAGG	60	
Db	1	TTTTTTTTTTTTTTTTTTCAAGTCCAAAGACATTTTTTTTTTTTTTTTTTTTATGATCAAGG	60	
Qy	61	ATTATTAAAGTCATACATGCAAAACATCTGCTAACTGCATAGCAAAAGATCAATGTAA	120	
Db	61	ATTATTAAAGTCATACATGCAAAACATCTGCTAACTGCATAGCAAAAGATCAATGTAA	120	
Qy	121	AAACACATCCCAAACTCTGCAACTGTCAAATGAAAAAAGTTTGTCTAGTGTGCAAAAGGC	180	
Db	121	AAACACATCCCAAACTCTGCAACTGTCAAATGAAAAAAGTTTGTCTAGTGTGCAAAAGGC	180	
Qy	181	CCAAACACTGTGTTCTTGCCAGTGAAGTTAGTTGTACAGAACGGCGTTAGCACTACGCGTT	240	
Db	181	CCAAACACTGTGTTCTTGCCAGTGAAGTTAGTTGTACAGAACGGCGTTAGCACTACGCGTT	240	
Qy	241	GACAGAACCTCAGACGCCAAAGGTPACCGGAAGCATGTGTGCGGTGGGTGAGGTCTAGA	300	
Db	241	GACAGAACCTCAGACGCCAAAGGTPACCGGAAGCATGTGTGCGGTGGGTGAGGTCTAGA	300	
Qy	301	GGGGCGGCATCAATCATGACAGTGTGTGCTCTGCGAAGACAGTGTGTTTCAGAA	360	
Db	301	GGGGCGGCATCAATCATGACAGTGTGTGCTCTGCGAAGACAGTGTGTTTCAGAA	360	
Qy	361	TATCTAAATAATAGTTTAAAAAACCTGTAAGCCGACACAGTGAATTTCTACACCAAGTTACTA	420	
Db	361	TATCTAAATAATAGTTTAAAAAACCTGTAAGCCGACACAGTGAATTTCTACACCAAGTTACTA	420	
Qy	421	GAAACGAAAGGAAGCACTAGTCAGCTGAGTAAAGGAAGGTGAAACAGGACGACATTC	480	
Db	421	GAAACGAAAGGAAGCACTAGTCAGCTGAGTAAAGGAAGGTGAAACAGGACGACATTC	480	
Qy	481	TACTATCTACCAAAAAAATCTCCGAATGCATTATCAGAAAGATCTTATAGTACAGGTGAG	540	
Db	481	TACTATCTACCAAAAAAATCTCCGAATGCATTATCAGAAAGATCTTATAGTACAGGTGAG	540	
Qy	541	ACATAATGCTCGTTTAAAGAGGGGGTCTTAAAGAAAGCACTTGCTGAAGTTAGCAATGTG	600	
Db	541	ACATAATGCTCGTTTAAAGAGGGGGTCTTAAAGAAAGCACTTGCTGAAGTTAGCAATGTG	600	
Qy	601	AGGATGGCCAGTTTAAATATGCACTCAACGCCCCATCTGGGAGGACAGCAGGGGGAG	660	
Db	601	AGGATGGCCAGTTTAAATATGCACTCAACGCCCCATCTGGGAGGACAGCAGGGGGAG	660	
Qy	661	GGGGCTCAAGAGAGACACTGATAGATCGGCCATTTGTCACTACTGTTTCACAGAAAT	720	
Db	661	GGGGCTCAAGAGAGACACTGATAGATCGGCCATTTGTCACTACTGTTTCACAGAAAT	720	
Qy	721	TAAACGGTTAAAAAGCTTTACCGGTGACACTTTTATTCAGTTGAAATTAATCTCCATGACAT	780	
Db	721	TAAACGGTTAAAAAGCTTTACCGGTGACACTTTTATTCAGTTGAAATTAATCTCCATGACAT	780	
Qy	781	GTAGTGTAAATTAATCTCTACTTCATATTAGTCAAAATACTGCTCTCTCTCTTGTATGAC	840	
Db	781	GTAGTGTAAATTAATCTCTACTTCATATTAGTCAAAATACTGCTCTCTCTCTTGTATGAC	840	
Qy	841	GTCTGTGTTTTCACACACTCCACCCAGCACACCCACGACTAGGAACAGAAATCTTCGTTAGA	900	

Db	841	GTGCTGTTTTCACACACTCCACCCAGCACACCCACGACTAGGAACAGAAATACTCTCGTTTGA	900
Qy	901	GGCAACACACAGGAGCCAGAGTCTGTTCAAAGCCCTGCAGAACGCCGTGCTGGTATTTTA	960
Db	901	GGCAACACACAGGAGCCAGAGTCTGTTCAAAGCCCTGCAGAACGCCGTGCTGGTATTTTA	960
Qy	961	GAGNACTCACTATCAAAATCAAAAGCAGAGCTGTTACACCCATCGTCAGCTACAGTACAA	1020
Db	961	GAGNACTCACTATCAAAATCAAAAGCAGAGCTGTTACACCCATCGTCAGCTACAGTACAA	1020
Qy	1021	AGTTACGCTAATGAGCATGGGCTGATTAAGTTACAGGTCGGTTACATGCACGGTGTCTATTA	1080
Db	1021	AGTTACGCTAATGAGCATGGGCTGATTAAGTTACAGGTCGGTTACATGCACGGTGTCTATTA	1080
Qy	1081	AGGAGGCTGTGCTGTGTCAACGGTCTGGGAGCTACGGGAGGCTCTGCACCCCTGAGCCC	1140
Db	1081	AGGAGGCTGTGCTGTGTCAACGGTCTGGGAGCTACGGGAGGCTCTGCACCCCTGAGCCC	1140
Qy	1141	AGAAGCTGCAGTCTTCTTTAAGGCAAAAGTCTCTCAACAGCTTAGTGTCTACGTGTTCTCA	1200
Db	1141	AGAAGCTGCAGTCTTCTTTAAGGCAAAAGTCTCTCAACAGCTTAGTGTCTACGTGTTCTCA	1200
Qy	1201	GCACAAACGCAACTTAGTGTCAAGGTAATTTTGGCAATTTCTAATCTCAGCAAGTAATGGG	1260
Db	1201	GCACAAACGCAACTTAGTGTCAAGGTAATTTTGGCAATTTCTAATCTCAGCAAGTAATGGG	1260
Qy	1261	GATTTT 1266	
Db	1261	GATTTT 1266	
RESULT 2			
AC106124/c			
LOCUS			
DEFINITION			
Rattus norvegicus clone CH230-119511, *** SEQUENCING IN PROGRESS			
***, 11 unordered pieces.			
ACCESSION			
AC106124			
VERSION			
AC106124.6 GI:30521557			
KEYWORDS			
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.			
SOURCE			
Rattus norvegicus			
ORGANISM			
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;			
Rattus.			
REFERENCE			
1. (bases 1 to 217700)			
Muzny D,Marie, Metzker M, Lee, Abramzon S, Adams C, Alder, J.,			
Allan, C, Allen H, Albrechts S, Amin A, Anguiano D,			
AUTHORS			

RESULT 2

AC106124/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

AC106124 217700 bp DNA linear HTG 10-MAY-2003

Rattus norvegicus clone CH230-119E11, *** SEQUENCING IN PROGRESS

***, 11 unordered pieces.

AC106124

AC106124.6 GI:30521557

HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1. (bases 1 to 217700)

Murny,D.Marie, Mettsker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalto,K., Blair,J., Blankenburg,K., Blyth,P.F., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,J., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregziorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Hagland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hughes,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idelbirt,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jollivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorenshewa,L., Loulsegged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,F., Martin,K., Martin,R., Martinez,E., Mawhinney,S., McLeod,M.P., McNelliff,Z., Meenen,F.,

Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nait, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwankwelu, O., Okwuonu, G., Olunpunsagbon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quirroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shwartsbeyn, A., Sisson, I., Sitter, C. D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steidle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Veta, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Unpublished
2 (bases 1 to 217700)
Worley, K.C.
Direct Submission
Submitted (12-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 217700)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On May 10, 2003 this sequence version replaced gi:25094619.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GHLL
Center clone name: CH230-119E11
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 185725 bases at least Q40
Consensus quality: 190140 bases at least Q30
Consensus quality: 192842 bases at least Q20
Estimated insert size: 194758; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will

* be preserved.
* 1 109797: contig of 109797 bp in length
* 109798 109897: gap of unknown length
* 109898 166934: contig of 57037 bp in length
* 166935 167034: gap of unknown length
* 167035 176384: contig of 9350 bp in length
* 176385 176484: gap of unknown length
* 176485 206327: contig of 29843 bp in length
* 206328 206428: gap of unknown length
* 206429 207461: contig of 1034 bp in length
* 207462 207561: gap of unknown length
* 207562 208607: contig of 1046 bp in length
* 208608 208707: gap of unknown length
* 208708 209908: contig of 1201 bp in length
* 209909 210008: gap of unknown length
* 210009 211583: contig of 1575 bp in length
* 211584 211683: gap of unknown length
* 211684 213749: contig of 1966 bp in length
* 213750 215115: contig of 1366 bp in length
* 215116 215215: gap of unknown length
* 215216 217700: contig of 2485 bp in length.
* Location/Qualifiers
* 1. 217700
* /organism="Rattus norvegicus"
* /mol_type="genomic DNA"
* /db_xref="taxon:10116"
* /clone="CH230-119E11"
* BASE COUNT 51384 a 45640 c 45954 g 51476 t 23246 others
* ORIGIN
* Query Match 97.0%; Score 1227.4; DB 2; Length 217700;
* Best Local Similarity 98.7%; Pred. No. 1.6e-312;
* Matches 1237; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
* QY 14 TTTTCAAGTTCCTCAAGACATTTTTTTTTTTTTTTTATGATCAAGATTTATTAAGTCA 73
* Db 64694 TTGCAAGTTCCTCAAGACATTTTTTTTTTTTTTTTATGATCAAGATTTATTAAGTCA 64635
* QY 74 TACATGCAAAACATACCTGCTAACTGCTATGCAAAAGATCAATGTAAACACCTCCACAA 133
* Db 64634 TACATGCAAAACATACCTGCTAACTGCTATGCAAAAGATCAATGTAAACACCTCCACAA 64575
* QY 134 TTTGCAACTGTCAATTCGAAAGAGTTTGTCTAGTGTGCGAAAGGCCCACTGTGTT 193
* Db 64574 TTTGTAACGTGCTCAATTCGAAAGAGTTTGTCTAGTGTGCGAAAGGCCCACTGTGTT 64515
* QY 194 CTTGCCAGTCAAGTAGTGTGTACGAAAGCGGTTAGCACTAGCGCTTGACAAACCTCAC 253
* Db 64514 CTTGCCAGTCAAGTAGTGTGTACGAAAGCGGTTAGCACTAGCGCTTGACAAACCTCAC 64455
* QY 254 AGACCCAAAGGTACCGGAAGCATGTGTCGCGTGGGTAGGTCTAGAGGGGGGGGATCA 313
* Db 64454 AGACCCAAAGGTACCGGAAGCATGTGTCGCGTGGGTAGGTCTAGAGGGGGGGGATCA 64395
* QY 314 ATCAGATGACAGTGTGGTACTCTGGCAACAGACAGTGTATTCAGAAATATCTAAATAGT 373
* Db 64394 ATCAGATGACAGTGTGGTACTCTGGCAACAGACAGTGTATTCAGAAATATCTAAATAGT 64335
* QY 374 TTAACAACTGTAAGCGCGGACGACCGTGTATTTACACCCAGTTACTAGAAACGAAAGGA 433
* Db 64334 TTAACAACTGTAAGCGCGGACGACCGTGTATTTACACCCAGTTACTAGAAACGAAAGGA 64275
* QY 434 AGCACTAGTCAAGTGTAGTAAAGGAGGTGAACAGACGACGACCTTCTACTTCTACCAA 493
* Db 64274 AGCACTAGTCAAGTGTAGTAAAGGAGGTGAACAGACGACGACCTTCTACTTCTACCAA 64215
* QY 494 AAAAACTCCGAATGCATTTATCAGAAAGATCTTATAGTACAGGTACAGATATTTGCTCGT 553
* Db 64214 AAAAACTCCGAATGCATTTATCAGAAAGATCTTATAGTACAGGTACAGATATTTGCTCGT 64155
* QY 554 TAAGAGGGGGTCTTAAGAAAGCACTTGCTGAAGTTAGCAACTGTGAGGATGGCCAGTT 613

Db	64154	TAAGAGGGGTCCTTAAGAAAGACATTGCTAGTACCACTGTGAGGTGGCAGTT	64095	Query Match	80.3%;	Score 1017;	DB 6;	Length 1017;
Qy	614	TAAATATGACTCAACGCCCATCTGGGAGGGACAGCAGGGGGAGGGGGCTCAAGAG	673	Best Local Similarity	100.0%;	Pred. No. 3.3e-257;		
Db	64094	TAAATATGACTCAACGCCCATCTGGGAGGGACAGCAGGGGGAGGGGGCTCAAGAG	64035	Matches 1017;	Conservative	0;	Mismatches	0;
Qy	674	AGACTGATAGATCGGCCATTGTGTCATCTACTCTTTCACAGAAATTAACCGTTTAAAA	733				Indels	0;
Db	64034	AGACTGATAGATCGGCCATTGTGTCATCTACTCTTTCACAGAAATTAACCGTTTAAAA	63975				Gaps	0;
Qy	734	GCTTTACCGCTGACACTTTTATTCAGTTGCAATTAATCTCCATGACAAATGATGTAATTA	793					
Db	63974	GCTTTACCGCTGACACTTTTATTCAGTTGCAATTAATCTCCATGACAAATGATGTAATTA	63915					
Qy	794	ATCTCTACTTCATATTAAGTCAAAATACTGCTGTCTCTCTTTCATGACGCTGTTTCA	853					
Db	63914	ATCTCTACTTCATATTAAGTCAAAATACTGCTGTCTCTCTTTCATGACGCTGTTTCA	63855					
Qy	854	CACTCCACCCAGCACACCCACACTAGGAACAGATACCTTCGTAGAGGCAACACAGGAG	913					
Db	63854	CACTCCACCCAGCACACCCACACTAGGAACAGATACCTTCGTAGAGGCAACACAGGAG	63795					
Qy	914	CCAGAGTCTGTTCAAGGCTCGAAGCGCGTCACTGCTGCTATTTTACAGAACTCACTAT	973					
Db	63794	CCAGAGTCTGTTCAAGGCTCGAAGCGCGTCACTGCTGCTATTTTACAGAACTCACTAT	63735					
Qy	974	GAAATCAAGAGCAGAGCTGTTACACCCATCGTACGTCACAGTACAGTAAATGA	1033					
Db	63734	GAAATCAAGAGCAGAGCTGTTACACCCATCGTACGTCACAGTACAGTAAATGA	63675					
Qy	1034	GCATGGGCTGATTAAGTTACAGTGCCTTACATGGCAGCGTGTCAATTAAGGAGCTGTCT	1093					
Db	63674	GCATGGGCTGATTAAGTTACAGTGCCTTACATGGCAGCGTGTCAATTAAGGAGCTGTCT	63615					
Qy	1094	GTCTCACGCTCTGGAGCTACGGAGGCTCTGACCCCTCGACCCCTGAGCCAGAACTGCAGTC	1153					
Db	63614	GTCTCACGCTCTGGAGCTACGGAGGCTCTGACCCCTCGACCCCTGAGCCAGAACTGCAGTC	63555					
Qy	1154	TTCTTAAGCAAAAGTCTCTCAACAGCTTAGTCTTACGTTCTAGCACAACCGCAACT	1213					
Db	63554	TTCTTAAGCAAAAGTCTCTCAACAGCTTAGTCTTACGTTCTAGCACAACCGCAACT	63495					
Qy	1214	TAGTTCACAGGATTTTGGCAATTTTGGCAATTTTGGCAATTTTGGCAATTTTGGCAATTTT	1266					
Db	63494	TAGTTCACAGGATTTTGGCAATTTTGGCAATTTTGGCAATTTTGGCAATTTTGGCAATTTT	63442					
RESULT 3								
AX163751/c								
LOCUS								
DEFINITION								
Sequence 15 from Patent WO0138579.								
ACCESSION								
AX163751								
VERSION								
AX163751.1								
KEYWORDS								
Rattus norvegicus (Norway rat)								
SOURCE								
ORGANISM								
Rattus norvegicus								
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;								
Rattus.								
REFERENCE								
1								
Gould-Rothberg B.E., Dippio V.A., Ramsehl T.M. and Gerwein R.W.								
METHOD OF IDENTIFYING TOXIC AGENTS USING NSAID-INDUCED DIFFERENTIAL								
GENE EXPRESSION IN LIVER								
Patent: WO 0138579-A 15 31-MAY-2001;								
Curagen Corporation (US)								
LOCATION/QUALIFIERS								
1. 1017								
/organism="Rattus norvegicus"								
/mol_type="genomic DNA"								
/db_xref="taxon:10116"								
FEATURES								
source								
BASE COUNT								
245 a								
ORIGIN								
245 c								
311 t								

1543	Db	CGTGCACCTTTTATT CAGTTGAATTA CTCAATGACAAATAGTGTAAAGTAATCTCTAC	1484
802	Qy	TTCTAATTAGTCAAAATACGTCTGTCTCTCTTTTGATGACGTCGTGTGTTTCACACACTCCAC	861
1483	Db	TTCTAATTAGTCAAAATACGTCTGTCTCTTTTGATCACTGTCTGTGTTT--CACACTCCAC	1426
862	Qy	CCAGCACACCCACGACTAGGAAACAGAATACCTCGTTAGAGGCAACACAGAGGCCAGAGTT	921
1425	Db	CCAGCACACCCACAACTAGGAAACAGAATACCTCATTAGAGGCAACACAGGAACGAGAGTT	1366
922	Qy	CTGTTCAAAGCCTGCAGAACCCGTCAGCTGGTATATTTAGAGAACTCACACTATGAATCAA	981
1365	Db	CTGTTCAAAGCTGCAAAAGCTAGTCGGCTGGCTTTTAGAAAACCTCACTATGAAATCAA	1306
982	Qy	AGAGCAGAGCTGTTACACCCATC--GTGACGTACAGTACAAAAGTTACGTAATGAGCATG	1038
1305	Db	AGAGCTGAGCTGTCACTCATCACTGTGACGTACAGTACAAAGTTACGTAATGACATG	1246
1039	Qy	GGCTGATAAGTTACAGGTGCGTTACATGGCAGCGGTGTCATTAAAGAGGCTGTGCTGTGTC	1098
1245	Db	GGCTGATAAGTTACAGGTGCGTTACATGGCAACGCTGTCAATAAGAGGAGGCTGTGCTGTGTC	1186
1099	Qy	ACAGGCTCTGGAGCTACGGGAGGCTCTGACCCCTGAGGCCCAAGAGCTCGACCTCTCTTT	1158
1185	Db	ACAGGCTCTGGGAACCTCGGAAGGCTCTGACCCCTGGCTCCCAAGAGCTCGGCTCTCTTT	1126
1159	Qy	AAGGACAAA--GTCTCTCAACAGCTTAGTGCTTTACGTGTCTTCAGCACAAACGCACTTA	1215
1125	Db	AGCAACAGAAGTCTCTCAAGAGAGTTTAGTGCTTTAGTGTTCTTCAGCACAAACGCACTTA	1066
1216	Qy	GTTCACAAGGTATTTTGGCAATCTTAAATCTGAGCAAGAATAGGGGATTTTT	1266
1065	Db	GTTTCAGAGGTATTTTGGCAATTTTAAATCTGAGCAAGAATAGGGGATTTTT	1015

RESULT 5	BC051053	2281 bp	linear	ROD 14-APR-2003
LOCUS	Mus musculus, clone MGC:58966		mrna	IMAGE:5038182, mRNA, complete cds.
DEFINITION	BC051053			
ACCESSION	BC051053.1	GI:29835221		
VERSION				
KEYWORDS	MGC.			
SOURCE	Mus musculus (house mouse)			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus;			
TITLE	1 (bases 1 to 2281)			
JOURNAL	Strausberg, R.			
	Submitted (11-APR-2003) National Institutes of Health, Mammalian			
	Gene Collection (MGC), Cancer Genomics Office, National Cancer			
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,			
	USA			
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov			
COMMENT	Contact: MGC help desk			
	Email: cgabps-remail.nih.gov			
	Tissue Procurement: Gilbert Smith, Ph.D.			
	cDNA Library Preparation: Life Technologies, Inc.			
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)			
	DNA Sequencing by: National Institutes of Health Intramural			
	Sequencing Center (NISC),			
	Gaithersburg, Maryland;			
	Web site: http://www.nisc.nih.gov/			
	Contact: nisc_mgc@nigri.nih.gov			
	Akhtar, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,			
	Blakesley, R.W., Bouffard, G.G., Green, K., Brinkley, C., Brooks, S.,			
	Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghghi, P.,			
	Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,			
	Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,			
	McDowell, J., Pearson, R., Stantirpop, S., Thomas, P.J., Touchman, J.W.,			
	Youngeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,			
	Young, A., Zhang, L.-H. and Green, E.D.			

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAK file: 108 Row: f Column: 9
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Genomescan gene prediction, Similarity but not identity to protein.

```

FEATURES             location/Qualifiers
source
i. .2281
/organism="Mus musculus"
/mol type="mRNA"
/strain="CZECH II"
/db xref="taxon:10090"
/clons="MGC:58966 IMAGE:5038182"
/tissue_type="Mammary tumor metastatized to lung. Tumor
arose spontaneously from a senescent normal mammary
(clonal) outgrowth infected with the virus MMIV."
/clone_lib="NCI_CGAP_Lu29"
/lab host="DH10B"
/notes="Vector: pCMV-SPORT6"
197 . 775
/codon_start=1
/product="Unknown (protein for MGC:58966)"
/protein_id="AAHS1053.1"
/db xref="GI:29835222"
/translations="MOAIOKVGVGDGAVGKTCLLIYTNNAFGEYIPTVDNYSANV
HVCQKPNVLGKLTAGQEDYDRPLSPYQDTFLFCFLVSPASENVRKAYPEYR
MHCNPNTILVGLTKDLRDDKDTIEKLKKELTPITYPQGLAMAKEIGAVKYLECSAL
TQRLGKTVDEAIRAVLCPPPPKKRKKCLLL"
568 a 569 c 504 g 640 t

BASE COUNT
ORIGIN
Query Match
68.6%; Score 968.2; DB 10; length 2281;

```

```
QY 550 TCCTTAAGAGGGGGTCTTAAGAAAGACATTTGCTAAGTTAGCAACTGTGAGGATGGCC 609
Db 1744 TCCTTAAGAA-GGGGTTCTTAAGAAAGACATTTGCTAAGTTAG-ACCTGTGAGGATGGCC 1687
QY 610 AGTTTAAATATGACTCAAGCCCAATCTGGGAGGAGGACAGAGGGGGAGGGGGCTCA 669
Db 1686 AGTTTAAATATGACTCAAGCCCAATCTGGGAGGAGGACAGAGGGGGAGGGGGCTCA 1627
QY 670 AG-----AGAGACACTGATAAGATCGGCCATTTGTTCATCTACTGTTTGACAGA 717
Db 1626 CGAAGGGATGCTCAAGAGACACTGATAAGATTTGGCCATTTGTTCATCTACTGTTTGACAGA 1567
QY 718 AATTAAACCTTTAAAGACTTTACCGGTGACACTTTTATTCAGTTGAAATTAATCTCCATGTAC 777
Db 1566 AATTAAACCTTTAAAGACTTTACCGGTGACACTTTTATTCAGTTGAAATTAATCTCCATGTAC 1507
QY 778 ATGTAGTGAATTAATCTACTTCATATTAATAGTCAAAATATCTGTCTCTCTTTGAT 837
Db 1506 AATGTAGTGAATTAATCTACTTCATATTAATAGTCAAAATATCTGTCTCTCTTTGAT 1447
QY 838 GACGTGCGTGTTCACACACTCCACCCAGCACACCCAGCAGTACAGAACTACTTCGTT 897
Db 1446 CATGTCGTTT--CACACTCCACCCAGCACACCCAGCAGTACAGAACTACTTCGTT 1389
QY 898 AGAGCAACACAGGAGCGAGTCTCTGTTCAAGGCTGCAGAGCGGGTCAGCTGCTATT 957
Db 1388 AGAGCAACACAGGAGCGAGTCTCTGTTCAAGGCTGCAGAGCGGGTCAGCTGCTATT 1329
QY 958 TTAGAGACTCACTATGAATCAAGAGCAGAGCTGTTTACCCCATC---GTGAGCTACA 1014
Db 1328 TTAGAGAACTCACTATGAATCAAGAGCAGAGCTGTTTACCCCATC---GTGAGCTACA 1269
QY 1015 GTACAAAGTTACGTAATCAGCATGGCTGATAGTTTACAGTGGCTTACATGGGAGCGTG 1074
Db 1268 GTACAAAGTTACGTAATCAGCATGGCTGATAGTTTACAGTGGCTTACATGGGAGCGTG 1209
QY 1075 TCATTAAGAGCGTGTGCTGTGTACAGCGTCTGGGAGCTACGGGAGGCTTCGACCCCT 1134
Db 1208 TCATTAAGAGCGTGTGCTGTGTACAGCGTCTGGGAGCTACGGGAGGCTTCGACCCCT 1149
QY 1135 GAGCCAGAGCTGAGTCTTTTAAGGAGAAA---GTCTCTCAACAGCTTAGTGCTTAC 1191
Db 1148 GACTCCAGAGCTGAGTCTTTTAGCAACAGAGAGTCTCTCAAGAGAGTTTAGTGCTTAC 1089
QY 1192 GTGTTCTCAGCACACCACTAGTTCACAGAGTATTTTGGCAATCTTAACTGTGACGA 1251
Db 1088 GTGTTCTCAGCACACCACTAGTTCAGAGAGTATTTTGGCAATTTTAACTGTGACGA 1029
QY 1252 AGAATAGGAGATTT 1266
Db 1028 AGAATAGGAGATTT 1014

RESULT 6
AC068493/c
LOCUS
DEFINITION
AC068493
ACCESSION
AC068493.10 GI:15148081
VERSION
HTG; HTGS PHASE1; HTGS DRAFT.
KEYWORDS
Mus musculus (house mouse)
SOURCE
Mus musculus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 269081)
REFERENCE
AUTHORS
Montgomery,K.T., Grills,G., Han,J., Lee,E., Long,J., Pomerantz,R.,
Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,
Gordon,M., Goltz,J.S. and Kucherlapati,R.
TITLE
High Throughput Mouse Sequencing
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 269081)
AUTHORS
Montgomery,K.T., Grills,G., Han,J., Lee,E., Long,J., Pomerantz,R.,
```

Ioshikhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A.,
Gordon, M., Goltz, J.S. and Kucherlapati, R.
Direct Submission
Submitted (03-MAY-2000) Department of Molecular Genetics, Albert
Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
Bronx, NY 10461, USA
On Aug 11, 2001 this sequence version replaced gi:14993654.
-----Genome Center
Center: Harvard Partners Genome Center
Center Code: HPGC
Web site: <http://www.hpcgg.org/Sequence/mouse.html>
Contact: hpgcmedel.mgh.harvard.edu
-----Summary Statistics
Center project name: ABN
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 100%
*Consensus quality: 256302 at least Q20
*Consensus quality: 254124 at least Q30
*Consensus quality: 250276 at least Q40
Estimated insert size: agarose-PP - N/A
**Estimated insert size: 268581 - sum-of-contigs
Quality coverage: agarose-PP - N/A
Quality coverage: 6.2 x in Q20 bases; sum-of-contigs estimation

* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
1 51730: contig of 51730 bp in length
51731 51750: gap of unknown length
51751 82423: contig of 30673 bp in length
82424 82443: gap of unknown length
82444 118158: contig of 35715 bp in length
118159 118178: gap of unknown length
118179 139441: contig of 21263 bp in length
139442 139461: gap of unknown length
139462 160638: contig of 21177 bp in length
160639 160658: gap of unknown length
160659 177611: contig of 16953 bp in length
177612 177631: gap of unknown length
177632 190991: contig of 13360 bp in length
190992 191011: gap of unknown length
191012 200128: contig of 9117 bp in length
200129 200148: gap of unknown length
200149 211882: contig of 11734 bp in length
211883 211902: gap of unknown length
211903 220773: contig of 8871 bp in length
220774 220793: gap of unknown length
220794 229643: contig of 8850 bp in length
229644 229663: gap of unknown length
229664 238339: contig of 8676 bp in length
238340 244848: gap of unknown length
244849 244868: gap of unknown length
244869 249725: contig of 4857 bp in length
249726 249746: gap of unknown length
249747 252524: contig of 2778 bp in length
252525 252544: gap of unknown length
252545 257520: contig of 4977 bp in length
257521 257541: gap of unknown length
257542 260397: contig of 2856 bp in length
260398 260416: gap of unknown length
260417 262710: contig of 2293 bp in length
262711 262729: gap of unknown length
262730 263934: contig of 1205 bp in length
263935 263954: gap of unknown length
263955 264196: contig of 242 bp in length
264197 264216: gap of unknown length
264217 264629: contig of 413 bp in length
```

```
* 264630 264649: gap of unknown length
* 264650 contig of 959 bp in length
* 265608 contig of unknown length
* 265609 contig of unknown length
* 265628: gap of unknown length
* 266335: contig of 1307 bp in length
* 266936: gap of unknown length
* 266955: gap of unknown length
* 268335: contig of 1380 bp in length
* 268355: gap of unknown length
* 268356: gap of unknown length
* 268356: contig of 194 bp in length
* 268356: gap of unknown length
* 268550: gap of unknown length
* 268570 269081: contig of 512 bp in length.
```

FEATURES

Source

```
1. 269081
   /organism="Mus musculus"
   /mol_type="genomic DNA"
   /strain="C57BL6/J"
   /db_xref="taxon:10090"
   /clone="RP23-76K1"
   /sex="male"
1. 51730
```

```
misc_feature 1. 51730
misc_feature /note="assembly_name:Contig138"
misc_feature 51751..82423
misc_feature /note="assembly_name:Contig137"
misc_feature 82444..118158
misc_feature /note="assembly_name:Contig136"
misc_feature 118179..139441
misc_feature /note="assembly_name:Contig135"
misc_feature 139462..160638
misc_feature /note="assembly_name:Contig134"
misc_feature 160659..177611
misc_feature /note="assembly_name:Contig133"
misc_feature 177632..190991
misc_feature /note="assembly_name:Contig132"
misc_feature 191012..200128
misc_feature /note="assembly_name:Contig131"
misc_feature 200149..211882
misc_feature /note="assembly_name:Contig130"
misc_feature 211903..220773
misc_feature /note="assembly_name:Contig129"
misc_feature 220794..229643
misc_feature /note="assembly_name:Contig128"
misc_feature 229664..238339
misc_feature /note="assembly_name:Contig127"
misc_feature 238360..244848
misc_feature /note="assembly_name:Contig126"
misc_feature 244869..249725
misc_feature /note="assembly_name:Contig125"
misc_feature 249746..252523
misc_feature /note="assembly_name:Contig124
clone end:SP6
vector_side:left"
misc_feature 252544..257520
misc_feature /note="assembly_name:Contig123"
misc_feature 257541..260396
misc_feature /note="assembly_name:Contig122"
misc_feature 260417..262709
misc_feature /note="assembly_name:Contig121"
misc_feature 262730..263934
misc_feature /note="assembly_name:Contig120"
misc_feature 263955..264196
misc_feature /note="assembly_name:Contig119"
misc_feature 264217..264629
misc_feature /note="assembly_name:Contig118"
misc_feature 264650..265608
misc_feature /note="assembly_name:Contig117"
misc_feature 265629..266935
misc_feature /note="assembly_name:Contig116"
misc_feature 266956..268335
misc_feature /note="assembly_name:Contig115"
misc_feature 268356..268549
misc_feature /note="assembly_name:Contig114"
misc_feature 268570..269081
misc_feature /note="assembly_name:Contig113"
BASE COUNT 69726 a 66027 c 64672 g 67724 t 932 others
```

ORIGIN

```
Query Match 68.1%; Score 862; DB 2; Length 269081;
Best Local Similarity 84.8%; Pred. No. 4.3e-216;
Matches 1101; Conservative 0; Mismatches 145; Indels 52; Gaps 10;

QY 14 TTTTCAAGTTCCAAAGACATTTTTTTTTTTTTTTTATGATTCAGAGATTTATTAAGTCA 73
Db 63492 TTGTCAAATTCAAAAACACAAGCATAACCACTGTATGATTCAGAGATTTATTAATCA 63433

QY 74 TACATGCAAAACATACCTACTGCTAATAGCAAAAGATCAATGTAAAAACATCCACAA 133
Db 63432 TACATGCAAAACATACCTGCTAATGCAATGAGCAAAAGATCAATGTAAAAACATCCACAA 63373

QY 134 TTCTGCAACTGTCAATTGAAAAAAGTTTCTTCTAGTGGTCGAAAGGCCCAACTGTGT 193
Db 63372 TTCTGCAACTGTCAATTGAAAAAAGTTTCTTCTAGTGGTTGAAAGGCCCAACTGCAIT 63313

QY 194 CTTGCCAGTGAGTTAGTTGTACAGAAACGGCGTTAGCAGTGGCTGTGACAGAACCTCAC 253
Db 63312 CTTGCCAGTGAGTTAGTTGTACAGAACAGCGTTAGCAGTGGCTGTGACAGAACCTCAC 63253

QY 254 AGACCCAAAGG-----TACCGGAAGCATGTGTCGGCTGGGTGA 292
Db 63252 AGACCCAAAGGACATCTCTAGCAGAGCCATGACAGAGAGCGTCTGTCCAGTGGCGGA 63193

QY 293 GGTCTAGAGGGCGCGCATCAATCACATGACAGTGTGGTACTCTGCGAAGACAGTGTATG 352
Db 63192 GGTCTAGA---GGCAGCATTTAGTCAATGACAGTGTGGTACTCTGTAGAGGTTTCATAG 63136

QY 353 TTTTCAAGTATCTAAATAGTTTAAAACTGTAAAGCCGCGACAGTGTCTTACACCC 412
Db 63135 TTTTCAAGTATCTAAATAGTTTAAAAACCGTAAAGCTGCAACACATGATTTCCACACCT 63076

QY 413 AGTTACTAGAAACGAGGAGGACACTAGT-CAGCTGAGTAAAGAGAGGTGAAAAACAGGA 471
Db 63075 AGTTGCTAGAAACGAGGAGGAGGACACTAGTCTAGCTCTGAGTAAAGAGGAGGAAACAGGA 63016

QY 472 ACGCACTTCTACTATCTACCAAAAAATCTCCGAA-----TGCAATATCAGAAAGATCTT 526
Db 63015 CGCACTTCTACTAGCTACCAAAAAATCTCCGAAATCTCGCTTATCAGAAAGATCTT 62956

QY 527 ATAGTACAGGTACAGATATTTGCTCGTTAAGAGGGGCTCTAAGAAAGACACTTCTGCTA 586
Db 62955 ATAGTACAGGTACAGATATTTGCTCGTTAAGAA- GGGGTTCTAAGAAAGACACTTCTGCTA 62897

QY 587 AGTTAGCAACTGTGAGGATGCGCAGTTTAAATATGGAATCAAGCCCATCTTGGGAGGG 646
Db 62896 AGTTAG-AACTGTGAGGATGCGCAGTTTAAATATGGAATCAAGCCCATCTTGGGAGGG 62838

QY 647 ACAGCAGGGGGAAGGGGGGCTCAAG-----AGAGACACTGATAGATCGGCGCA 694
Db 62837 ACAGCAGTGGAGGTTGGGGGGCGGAAGGATGCTCAAGAGACACTGATAGATCGGCGCA 62778

QY 695 TTTGTCATCTACTGTTTGACAGAAATTAACCGTTAAAGGCTTTACCGTGACACTTTTA 754
Db 62777 TTTGTCACCTACTGTTTGACAGAAATTAACCGTTAAAGGCTTTACCGTGACACTTTTA 62718

QY 755 TTCAGTTGAATTAATCTCCATGTACAACTGATGTGTAATAATCTCTACTTCTATTAAGTCA 814
Db 62717 TTCAGTTGAATTAATCTCCATGTACAACTGATGTGTAATAATCTCTACTTCTATTAAGTCA 62658

QY 815 AAATACTGTCTGTCTCTCTTTGATGACGTCGTTTTCACACTCCACCCAGCACACCCAC 874
Db 62657 AAATACTGTCTGTCTCTCTTTGATGATCATGTCGTGTTT--CACACTCCACCCAGCACACCCAC 62600

QY 875 GACTAGGAAACAGAAATCTTCTGTAGAGGCAACACAGGAGCGAGAGTCTGTTCAAAGCT 934
Db 62599 AACTAGGAAACAGAAATCTTCTATTAGAGGCAACACAGGAGCGAGAGTCTGTTCAAAGTCT 62540

QY 935 GCAGAAGCGGTGAGCTGGTATTTTAGAGAACTCACTATGAAATCAAAAGACAGAGTGT 994
Db 62539 GCAGAAGCTAGTCGGCTGGTCTTTTAGAGAACTCACTATGAAATCAAAAGAGCTGAGTGT 62480
```

```

QY 995 TACACCATC---GTGACGTACAGTACAAAGTACGTAATGAGCAGTGGCTGATAGTTA 1051
Db 62479 CACACTCATCTGCTGACGTACAGTACAAAGTACGTAATGAGCAGTGGCTGATAGTTA 62420
QY 1052 CAGGTGGCTTACATGCGCAGCGTGTGCTTAAAGAGGCTGTGCTGTGTCTACACGGTCTGGGA 1111
Db 62419 CAGGTGGCTTACATGCGCAGCGTGTGCTTAAAGAGGCTGTGCTGTGTCTACACGGTCTGGGA 62360
QY 1112 GCTAGGGAGGCTGACACCGCTGAGCCGAGAGCTGAGTCTTCTTAAAGACAAA---G 1168
Db 62359 ACTCTGAAAGGCTGTGACACCGCTGAGCCGAGAGCTGAGTCTTCTTAAAGACAAA---G 62300
QY 1169 TCTCTCAACAGCTTACGTCTTACGTGTCTCAGCAGACGCAACTTACGTCTCAGAGGTAT 1228
Db 62299 TCTCAAGAGGTTAGTGTCTTACGTGTCTCAGCAGACGCAACTTACGTCTCAGAGGTAT 62240
QY 1229 TTTGGCAATCTTAATCTGACGAAGATPAGGGGATTT 1266
Db 62239 TTTGGCAATTTTAATCTGACGAAGATPAGGGGATTT 62202

RESULT 7
BC050687/c
LOCUS
DEFINITION
  BC050687 2302 bp mRNA linear PRI 11-APR-2003
  Homo sapiens, ras-related C3 botulinum toxin substrate 1 (rho
  family, small GTP binding protein Rac1), clone MGC:60264
  IMAGE:6149377, mRNA, complete cds.
ACCESSION
  BC050687
VERSION
  BC050687.1 GI:29792301
KEYWORDS
  MGC.
SOURCE
  Homo sapiens (human)
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 2302)
  Direct Submission
  Submitted (08-APR-2003) National Institutes of Health, Mammalian
  Gene Collection (MGC), Cancer Genomics Office, National Cancer
  Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
  USA
  NIH-MGC Project URL: http://mgc.nci.nih.gov
  Contact: MGC help desk
  Email: cgaphs-x@mail.nih.gov
  Tissue Procurement: ATCC/DCTP/DTF
  cDNA Library Preparation: Life Technologies, Inc.
  DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
  Center, Stanford University School of Medicine, Stanford, CA 94305
  Web site: http://www-ehgc.stanford.edu
  Contact: (Dickson, Mark) mcd@paxil.stanford.edu
  Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
  R. M.
  Clone distribution: MGC clone distribution information can be found
  through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
  Series: IRAX Plate: 110 Row: K Column: 14
  This clone was selected for full length sequencing because it
  passed the following selection criteria: matched mRNA gi: 9845510.
  Location/Qualifiers
    1..2302
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="MGC:60264 IMAGE:6149377"
      /tissue_type="Skin, melanotic melanoma."
      /clone_lib="NIH MGC_72"
      /lab_host="DH10B"
      /note="Vector: pCMV-SPORT6"
      189..767
        /codon_start=1
        /product="ras-related C3 botulinum toxin substrate 1 (rho
  CDS

```

```

family, small GTP binding protein Rac1)
/protein_id="AAH50687.1"
/db_xref="GI:29792302"
/db_xref="LocusID:5879"
/translations="MOAIKVVVGDDGAVGKTCILSYITNAPGPEYITVFDNYSANV
NVDKPNVLGNDTAGDEDYDLRLPLSYFDVDFELICFSLVSPASFENVRKWPVEVR
HHCNTTILVGTKLDLRDDKDTIEKLEKKLTFTTYPQGLAMAKEIGAVKYLECSAL
TQRLKTVFPEAIRAVLCPFPVKKRKKLL"
BASE COUNT 617 a 524 c 498 g 663 t
ORIGIN
  Query Match 35.2%; Score 445.6; DB 9; Length 2302;
  Best Local Similarity 68.2%; Pred. No. 2e-106;
  Matches 874; Conservative 0; Mismatches 324; Indels 84; Gaps 15;
  QY 35 TTTTNTTTTTTTTTTATGATTCGAAGGATTTTAAAGTCATACATCGCAAAACATATGCTA 94
  Db 2302 TTTTNTTTTTTTTTTATGATTCGAAGGATTTTAAAGTCATACATCGCAAAACATATGCTA 2243
  QY 95 ACTGCATTAGCAAAAAGATCAATGTAAAAACACTCCCAANTTCTCAACTGTCAATTGAAA 154
  Db 2242 ATTGCATTAGCAAAAAGATCAATGTAAAAACACTCCCAANTTCTCAACTGTCAATTGAAA 2183
  QY 155 AAAGCTTTTCTAGTGTGCGAAGGCCCAACACTGTGTTCTTGCCAGTGAAGTTAGTTGT 214
  Db 2182 AAATCTGTTGTAGTGTGCGAAGGCCCAACACTGTGTTCTTGCCAGTGAAGTTAGTTGT 2123
  QY 215 ACAGACGGGGTTAGCACTAGCGC--TTGACAGAACCTTCAGACCCCAAGG-----264
  Db 2122 ACAGAACATGTCAGCACTAGCAGTTTACAGAACCTTCAGACCCCAAGGAAACATCAA 2063
  QY 265 -----TACCGAAGCATGTGTCGGGTGGTGAAGTCTAGAGGGGGCGGCAT 311
  Db 2062 TAGCAAAAGCGACTACAGAGGGCGTGTGTCGGGTGGCGAGGTAAAGA--GGGTCAAGTAT 2004
  QY 312 CAATCATACAGCAGTGTGTTGTTCTCTGGCAAGCAGTGAAG-----TT 354
  Db 2003 TGGTCAAGTCAGCAGTGTGCGGTAAATCTGGCAAGCAGTGAAGTTTAAAGAGGTTTCAAGTTT 1944
  QY 355 TCAGAAATATCTAAATAAGTTTAAAAAAGCTGAAAGCCGACAGCTGATTTCTACACCCAG 414
  Db 1943 AAGAAATATCTAAATAAGTTTAAAAAAGCTGAAAGCCGACAGCTGATTTTACACCTAG 1884
  QY 415 TTACTAGAAAACGAAGGGAAGCACTAGTCAAGTGAAGTAAAGGAAGGTGAAAAACAGGAACG 474
  Db 1883 TTACTAGAAAACCTAAGGAAAGCACTTATTAGTCTGAATA-----AAGTAAATGGAAG 1829
  QY 475 CACTTCTACTATCTACCAAAAAAATCTCCGAATGCATTATCAGAAAGATCTTATAGTACA 534
  Db 1828 CACTTTTACTAAATCGCAAAAAAACCTTCTAATGTCATTATCAGAAAGATTTTATAATACA 1769
  QY 535 GGTGACACATATTGCTCGTTAAGAAAGGGGTCTTAAAGAAAGACACTTGTAAAGTTAGCA 594
  Db 1768 AG-GAGCATATTGCTCAGTCAGAGGGGTTCTATAGAAAGACCTTCTAAGTTAGCG 1710
  QY 595 ACTGTGAGGATGCCAGTTTAAATATGAACTAACGCCCATCTGGGGAGGACAGCAGG 654
  Db 1709 ACTTAACAGAACCAACAGTTTAAAGATGAATTAATAGCCCAATTTGGGAGGATGGCAGG 1650
  QY 655 GGGAGGGGGCTCAGAGAGACACTGATAAGATCGGCCATTTGTCTACTCTGTTTGTAC 714
  Db 1649 TGTAAGAGAAAGGAAAGGCTTAAAGAAACATTTCTGTGATAATACCAACCTTTCTTTCATC 1590
  QY 715 AGAAATTAACCGTTAAAAAGCTTTTACCCGTGACACTTTTATTCAGTTGAA---TTACTCC 771
  Db 1589 ATCTACTGCTATTGACAGAAATTAACCTTTAGAGTTTTTACCGTGACACTTCTATTC 1530
  QY 772 ATGTACAATGTAGTGAATTAATCTCTACTTCATA-TTAGTCAAAATACTGTCTGTCTC 830
  Db 1529 TTGTACAATGTAGTGTA-----AATCTCCACTTCGTATTTTGTCAAAATACTGTCTTGTG 1474
  QY 831 CTTTGTATGACGTCGGTGTTCACACACTCCACCCAGCACCCACGACTAGGAACAGAA 890

```



```

Db      1473 CTTTGATCA-----CACACACCCACCCGCGACACACACAGCTA--AACAGAAATT 1426
Qy      891 CTTTCGTTAGAGCAACACAGGAGCCAGAGTCTTCTTTCAAAGCCTCGAGAGCGGTCAGC 950
      |||||
Db      1425 CTTTCATTAGAGGAATA-----GCAGTTCGTCTTCAAAATCTCCGCAAAAGCTGGTTCAGA 1372
Qy      951 TGGTATTTTAGAGAACTCACTATGAATCAAAAGACAGAGCTGTTACACCCATCGTGGAG 1010
Db      1371 AAATCGCTATGAATCAAAAGACTGATCCAAAGAGCTGAGCTGTACGCTCACTCCAT 1312
Qy      1011 TACAGTACAAAGTTACGTAATGAGATGGCTGATAAGTTACAGGTGCGCTTACATGGCAG 1070
Db      1311 TACAGTACAAATGTTATGTCGGACACAGGCTGCTAACTCACTGGTGAAGTTCAATGGCAA 1252
Qy      1071 COTGTCATTAAGAGAGCTGTGCTGTCTACAGGCTCTGG-----GAGCTAGGGAGGG 1123
Db      1251 CGCTTCATTCGGGAGGCTGTTCTGCTTTACGCATCTGAGAACTACATAGGAGAGCAAGTG 1192
Qy      1124 TCTGCACC--CCTGAGCCCAAGAGCTGACGTCTTCTTTAAGGACAAAGTCTCTCAACAGCTT 1182
Db      1191 TCTGCACCTCTTAACGTGCAAGAGCTACCGTCTTCTCAAGACGAAGTCTTTGCAAGTT 1132
Qy      1183 ACTGCTTAGCTGTCTTCAGCAACAC---GCAACTTAGTTTCAAGAGTATTTTGGCAATTC 1239
Db      1131 CAGTGTCTGGTGTCTTCGGCACACAACTGACGTGTAGTTTCAGAAGTATTTTGGCAATC 1072
Qy      1240 TTAATCTGACGAAGATAGGGG 1261
Db      1071 TTAATCTGAACAAGATGGGG 1050

RESULT 8
HSA132695/c
LOCUS      28567 bp      DNA      linear      PRI 09-NOV-2000
DEFINITION Homo sapiens rac1 gene.
ACCESSION  AJ132695
VERSION     AJ132695.5 GI:8574037
KEYWORDS   alternative splicing; Alu; AluJo; AluSg1; AluSp; AluSx; Alu-rich;
            CT-rich; MIR; rac1 gene; rac1 protein; Rac1b protein; repetitive
            sequence.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS     Matos,P., Skaug,J., Marques,B., Beck,S., Verissimo,F., Gespach,C.,
            Boavida,M.G., Scherer,S.W. and Jordan,P.
            Small GTPase Rac1: structure, localization, and expression of the
            human gene
            J Biol Chem 277 (3): 741-751 (2000)
JOURNAL     Biochem Biophys. Res. Commun. 277 (3), 741-751 (2000)
MEDLINE     20517245
PUBMED      11062023
REFERENCE   2 (bases 1 to 28567)
AUTHORS     Jordan,P.
TITLE       Direct Submission
JOURNAL     Submitted (03-FEB-1999) Jordan P., Centro de Genetica Humana,
            Laboratorio de Oncobiologia, Instituto Nacional de Saude 'Dr.
            Ricardo Jorge', Avenida Padre Cruz, 1649-016 Lisboa, PORTUGAL
COMMENT     On Jun 20, 2000 this sequence version replaced gi:7248282.
            Related sequence AF638561.
FEATURES
            Location/Qualifiers
            1..28567
               /organism="Homo sapiens"
               /mol_type="genomic DNA"
               /db_xref="taxon:9606"
               /chromosome="7"
               /map="7p22"
               /clone="H. NH0425A05, RPCI 11"
               /c109
            repeat_region 1..109
            repeat_region 125..422
            repeat_region /rpt_family="AluSg"
            repeat_region /rpt_family="AluSc"
            repeat_region 441..742

```

```

/rpt_family="AluSg"
743..28117
/gene="rac1"
743..1226
/gene="rac1"
911..983
/rpt_family="GC_rich"
936
/gene="rac1"
/note="alternative transcription start site"
1026
/gene="rac1"
/note="alternative transcription start site"
1037
/gene="rac1"
/note="alternative transcription start site"
1053
/gene="rac1"
/note="alternative transcription start site"
1106..1175
/rpt_family="GC_rich"
/rpt_family="GC_rich"
1227..1261,13701..13772,18413..18530,22804..22860,
24268..24330,26010..26169,26458..26588)
/gene="rac1"
/note="alternative"
/codon_start=1
/product="Rac1b protein"
/protein_id="CAA10733.6"
/db_xref="GI:8574039"
/db_xref="SPTREMBL:O95501"
/translation="MQAICVVVGGAVGKTCLLISYTTNAPFGEYIPTVFDNYSANV
MVDGKPNVLGMDTAGQEDYDLRLPLSPYQTVGYGKDTSRGKDKPIADVPLICFS
LVSPASFENVRKAWPEVREHCPNTPILVGTKLDRDDKDTIEKLKKLLTPTVPO
GLAMAKEGAVKYLECSALTQRLKTVFEDEAIRAVLCPPPVKKRKKCLLL"
Join(1227..1261,13701..13772,18413..18530,24268..24330,
26010..26169,26458..26588)
/gene="rac1"
/codon_start=1
/product="Rac1 protein"
/protein_id="CAB53579.5"
/db_xref="GI:8574038"
/translation="MQAICVVVGGAVGKTCLLISYTTNAPFGEYIPTVFDNYSANV
MVDGKPNVLGMDTAGQEDYDLRLPLSPYQTVGYGKDTSRGKDKPIADVPLICFS
LVSPASFENVRKAWPEVREHCPNTPILVGTKLDRDDKDTIEKLKKLLTPTVPO
GLAMAKEGAVKYLECSALTQRLKTVFEDEAIRAVLCPPPVKKRKKCLLL"
Join(1227..1261,13701..13772,18413..18530,24268..24330,
26010..26169,26458..26588)
/gene="rac1"
/codon_start=1
/product="Rac1 protein"
/protein_id="CAB53579.5"
/db_xref="GI:8574038"
/translation="MQAICVVVGGAVGKTCLLISYTTNAPFGEYIPTVFDNYSANV
MVDGKPNVLGMDTAGQEDYDLRLPLSPYQTVGYGKDTSRGKDKPIADVPLICFS
LVSPASFENVRKAWPEVREHCPNTPILVGTKLDRDDKDTIEKLKKLLTPTVPO
GLAMAKEGAVKYLECSALTQRLKTVFEDEAIRAVLCPPPVKKRKKCLLL"
<1227..1261
/gene="rac1"
/number=1
1262..13700
/gene="rac1"
/number=1
1390..1419
/rpt_family="GC_rich"
1752..1793
/rpt_family="FLAM"
1794..2005
/rpt_family="AluJo"
2140..2171
/rpt_family="Simple_repeat"
2666..2972
/rpt_family="AluSx"
2980..3267
/rpt_family="AluSx"
3268..3289
/rpt_family="(T)n"
3802..4102
/rpt_family="AluJb"
4122..4386
/rpt_family="AluSx"
4538..4719
/rpt_family="LTR33A"
4754..5053
/rpt_family="AluSg"

```


Db 27375 TTTTACCGGTGACACTTTTCCTTCCTGATACAACTAGTGTGA-----AATCTCCACTTCGT 27320

QY 807 A-TTGTGCAAAATAGTCTGTCTCTCTTTGATGACGTCGTGTTTTCACACACTCCACCCAG 865

Db 27319 ATTGTGTCAAAATAGTCTGTCTCTCTTTGATGAC-----CACACACCCACCCCGG 27270

QY 866 CACACCCAGCTAGGACAGAAATACTCTGTTAGAGGCAACACAGGACGAGTGTCTGT 925

Db 27269 CACACCCACAGTA--ACAGAAATCTTCTATTAGAGGAATA-----CGAGTCTGTTC 27218

QY 926 TCAAGCCTGCGAAGCCCGTCAGTCGTGTTATTTAGAGAACTCATATGAATCAAGAG 985

Db 27217 AAAATCTCGCAAAAGCTGTCAGAAACTCGCTATGAATCACAAGAGCTGATCCAAAG 27158

QY 986 CAGAGCTGTACACCCATCGTCAGCTACAGTACAGTCAAGTTACGTAATGAGCATGGCTGAT 1045

Db 27157 AGCTGAGCTGCTACGCTCTACTCCATTACAGTCAATGTTATGTGCGGAACAGTGTCTGT 27098

QY 1046 AAGTTACAGGTGCTTACATGGCAGCGTGTCTATTAGAGAGGCTGTCTGTGTACACAGGT 1105

Db 27097 AACTCACTGTTGAGTTCATGCAACGCTTCATTCGGGAGGCTGTTCTGTTTACGCATC 27038

QY 1106 CTGG-----GAGTACCGGAGGGTCTGCACC-CTGAGCCCAAGAGCTGCAGTCTTCT 1157

Db 27037 TGAGAACTACATAGGAGAGCAAGTGTCTGCACCTCCTAACTGCAGAAGCTACCGTCTTCT 26978

QY 1158 TAAGGCAAGTCTCTCAACAGTGTAGTGTCTACGTTGTTCTCAGCAACAC---GCACTT 1214

Db 26977 CAAAGCAGAGGTCTTTGCAAGTTCAGTCTCGGTGTTCTCGGCACAAATGCAAGTGT 26918

QY 1215 AGTTCAAGAGTATTTTGGCAATCTTAATCTCAGCAAGAAATAGGGG 1261

Db 26917 AGTTCAGAGGTATTTGGCACTCTTAATCTCAGCAAGATGGGG 26871

RESULT 9
AC009412/c

LOCUS 212827 bp DNA linear PRI 07-NOV-2001

DEFINITION Homo sapiens BAC clone RP11-425P5 from 7, complete sequence.

AC009412

ACCESSION AC009412.6 GI:14190769

VERSION HTG.

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 212827)
Bukaryofa; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Sulston, J.E. and Waterston, R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
8947074

REFERENCE 2 (bases 1 to 212827)
Hou, S., Maupin, R., Haakenson, W., Gregory, S. and Belter, E.
The sequence of Homo sapiens BAC clone RP11-425P5
Unpublished

REFERENCE 3 (bases 1 to 212827)
Waterston, R.H.
Direct Submission
Submitted (21-AUG-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 4 (bases 1 to 212827)
Waterston, R.H.
Direct Submission
Submitted (23-MAY-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 5 (bases 1 to 212827)
Waterston, R.
Direct Submission
Submitted (07-NOV-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

COMMENT

On May 23, 2001 this sequence version replaced gi:13431187.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
Center project name: H_NH0425P05

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send mailto:egreen@nhgri.nih.gov, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RP11-11 human BAC library was made from the blood of one male donor, as described by Googawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanesi, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is CTD-2195F2, 200 bp overlap. Actual start of this clone is at base position 1 of RP11-425P5; actual end is at base position 212633 of RP11-425P5.

FEATURES

source

1. 212827
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="7"
/map="7"

repeat_region

/clone="RP11-425P5"
1623..1696
/rpt_family="MIR"

repeat_region

1698..2006
/rpt_family="Alu"

repeat_region

2024..2337
/rpt_family="Alu"

repeat_region

3006..3093
/rpt_family="L1"

repeat_region

3453..3552
/rpt_family="L2"

repeat_region

4924..5046
/rpt_family="ERV1"

misc_feature

5318..5693
/note="similar to EST BF996284 (NID:gl2402607)"

repeat_region

5580..5845
/rpt_family="Alu"

repeat_region

5925..6277

HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.

FEATURES

source Location/Qualifiers
1..2315
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="BRACE2008968"
/tissue_type="cerebellum"
/clone_lib="BRACE2"
/note="cloning vector: pME18SFL3"
BASE COUNT 603 a 531 c 501 g 680 t
ORIGIN

Query Match 34.0%; Score 430.6; DB 9; Length 2315;
Best Local Similarity 67.8%; Pred. No. 1.8e-102; Indels 84; Gaps 15;
Matches 859; Conservative 0; Mismatches 324;

QY 50 ATGATTTCAGGATTTATTAAAGTCATACATGCAATGCAAAACATCTGCTAACTCGATTAGCAAAA 109
Db 2315 ATGATTTCAGGATTTATTAAAGTCATACATGCAATGCAAAACATCTGCTAACTCGATTAGCAAAA 2256
QY 110 GATCAATGTAAATAACACCTCCCAATTCGCAACTGTCATTTGAAATAAGTTTGTCTAGT 169
Db 2255 GATCAATGTAAATAACACCTCCCAATTCGCAACTGTCATTTGAAATAAGTTTGTCTAGT 2196
QY 170 GGTGAAAGGCCCAACACTGTCTTCTGCGAGTGTAGTTGTACAGAACGGCGGTAG 229
Db 2195 GGTGAAAGGCCCAACACTGTCTTCTGCGAGTGTAGTTGTACAGAACGGCGGTAG 2136
QY 230 CACTAGCGC - TTGACAGAACCTCACAGACCCCAAGG - - - - - TA 266
Db 2135 CACTAGCACAGTTTACAGAACCTCACAGACCCCAAGGACATCAATAGGCAAGCGACTA 2076
QY 267 CGGAAAGCATGTGTCGCGGTGGTGAGGTCTAGAGGGGGGGCGGATCAATCAGATGACGT 326
Db 2075 CAGGAGGCGGTGTCGCGGTGGTGAGGTCTAGAGGGGGGGCGGATCAATCAGATGACGT 2017
QY 327 GTTGGTACTCTGCGACAGACAGTGTG - - - - - TTTTCAGAAATATCTAAAA 369
Db 2016 GTGCGTAATCTGCGACAGACAGTGTGTTAAGAGGTTTCATAGTTTAAAGATTTATCTAAAA 1957
QY 370 TAGTTTAAAAAAGCTGTAAGCGCGCAGCAGTGTATTTCTACACCCAGTTTACTAGAAAAAGCA 429
Db 1956 TATTTTAAAAAAGCTGTAAGAGTGTGCAACACATGATTTTTCACACCTAGTTTACTAGAAAAAGCA 1897
QY 430 GGGAAAGCATGTGCTGAGTGTAAAGAGGTGAAACAGGACCGACCTTCTACTATCTA 489
Db 1896 GGAAGGACATTTAGTCTCTGAATA - - - - - AAGTAACATGGAAGGACCTTTTACTAATCG 1842
QY 490 CCAAAAAATCTCGAATGCATTTATAGTAAGATCTTATAGTACAGTCTAGACATATTTGC 549
Db 1841 AAAAAAAGACCTTCTAATGATTTATAGTAAGATCTTATAGTAAGATTTATAGTAAGATTTGC 1783
QY 550 TCGTTAAGAGGGGGTCTTAAAGAAAGACCTTCTGTAAGTTAGCAACTGTGAGGATGGCC 609
Db 1782 TCAGTCAGAGGGGGTCTTAAAGAAAGACCTTCTAAGTTAGGACCTTACTAAGTTAGGACCTTAAAGAAAGCA 1723
QY 610 AGTTTAAATATGGAATCAAGCCCGCATCTCGGGAGGACAGGAGGGGGAGGGGGCTCA 669
Db 1722 AGTTTAAAGATGAATTAATGCGCAATTTGGGAGGATGCGAGGTTTAAAGAAAGAGAA 1663
QY 670 AGAGACACATGATAAGATCGCCATTTGTCATCTACTGTTTGCAGAAATTAACCGTTA 729
Db 1662 AAGCTTAAAGAAACATTTCTGATATACAACTTTCTTCTCATCTACTGCTATTTGA 1603
QY 730 AAAAGCTTTTACCCGTGACACTTTTATTCAGTTGAA - - - - - TTTACTCATGTACAATGTAGT 786
Db 1602 CAGAAATTAACCTTTTATAGGTTTATACCCGTGACACTTTCTTCTTCTGTAATGTAGT 1543
QY 787 TAAATTAATCTCTACTTCTA - TTAGTCAAAATAGTCTGCTCTCTCTTGTATGATGCTGT 845

Db 175463 TGATAATACCAACCTTTCTTTCATCATCTAGTCAATTTGACAGAAATTAACCTTTTAGAG 175404
QY 750 TTTTATTTCAGTTGAA - - - - - TTACTCCATGACAAATAGTGTAAATTAATCTTCTACTTCAT 806
Db 175403 TTTTACCCGTGACACTTTCATCTCTTGTACAAATGAGTGTGTA - - - - - AATCTCCACTTGGT 175348
QY 807 A - TTAGTCAAAATATGCTGTCTCTCTTTGATGAGTGTGTTTTCACACACTCCACCCAG 865
Db 175347 ATTTTGTCAAAATATGCTGTCTCTCTTTGATGTA - - - - - CACACACCCACCCCGG 175298
QY 866 CACACCCAGCTAGGACAGAAATCTGTTAGAGGCAACACAGGACGACAGTCTTGT 925
Db 175297 CACACCCACAGCTA - - - - - AACAGAAATCTTCAATGAGGAATA - - - - - GCAGTTCTGTTC 175246
QY 926 TCAAGAGCTCGAAGACCGGTGAGTGTATTTTATAGAACTCACTATGAAATCAAGAG 985
Db 175245 AAAATCTCGCAAAAGCTGTGCAAAACCTGCTATGAATACAAAGACTGATCCAAAG 175186
QY 986 CAGAGCTGTTCACCCATCGTACAGTACAGTACAAAGTTACGTAATGAGCATGGGCTGAT 1045
Db 175185 AGCTGAGCTGCTACGCTCACTCCATACAGTACAAATGTTATGTCGGGAACACGTGCTGT 175126
QY 1046 AAGTTACAGTGGTTTACATGCGAGCGTGTCTATTAAGGAGGTGTCTGTGTACACAGGT 1105
Db 175125 AACTCACTGGTGAATTCATGGAACGCTTCAATTCGGAGGCTGTCTGTCTTTACGATC 175066
QY 1106 CTGG - - - - - GAGCTACGCGGAGGCTCTGCACC - CTTGAGCCCAAGAGTGTGCTTCT 1157
Db 175065 TGAGAACTACATAGGAGAGCAAGTGTCTGCACCTCTCACTGCAAGAGTACCGTCTTCT 175006
QY 1158 TTAGGCAAAAGTCTTCAACAGCTTAGTGTACGTTTCTTACGACACAC - - - - - GCAACT 1214
Db 175005 CAAAGACGAAGTCTTTGCAAAAGTTCAGTCTCGGTGTCTTCTCGGCAACAATGCAAGT 174946
QY 1215 AGTTCAAGAGTATTTTGGCAATCTTAATCTCAGCAAGAAATAGGGG 1261
Db 174945 AGTTCAAGAGTATTTTGGCAATCTTAATCTCAGCAAGAAATAGGGG 174899

RESULT 10

AK054993/c

LOCUS

DEFINITION

Homo sapiens cDNA FLJ30431, clone BRACE2008968, highly similar

to RAS-RELATED C3 BOTULINUM TOXIN SUBSTRATE 1.

AK054993

AK054993.1 GI:16549633

KEYWORDS

oligo capping; fis (full insert sequence).

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Eukaryota; Euthera; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1

Tashiro H., Yamazaki M., Watanabe K., Kumagata A., Itakura S.,

Fukushima Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,

Osuski T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,

Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,

Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,

Wagatsuna M., Murakawa K., Kanehori K., Takahashi-Fujii A.,

Oshima A., Sugiyama A., Kawakami B., Suzuki Y., Sugano S.,

Nagahara K., Masuko Y., Nagai K. and Sugiyama T.

NEDO human cDNA sequencing project

Unpublished

2 (bases 1 to 2315)

Isogai T., Otsuki T. and Sugiyama T.

Direct Submission

Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,

Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan

(E-mail: genomics@helix.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)

NEDO human cDNA sequencing project supported by Ministry of

Economy, Trade and Industry of Japan; cDNA full insert sequencing:

Research Association for Biotechnology (RAB); cDNA library

construction: Helix Research Institute (HRI) (supported by Japan

key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,

```

Db      1542 TA-----AATCTCCACTTCGTATTTTGTCAAATACTGCTTTGTCTCTTGATCA----- 1493
QY      846 GTTTCACACTCCACCAGCACACCCAGCATAGAAACAATATCTTGGTTAGAGGCAA 905
Db      1492 ----CACACACCCACCCGGCACACCCACAGCTA--AACAGAAATCTTTCATTAGAGGAAA 1439
QY      906 CACAGGAGCCAGAGTTCGTTCAAAGCCTGCAGAGCCGGTCCAGCTGGTATTTTAGAAA 965
Db      1438 TAG-----CAGTTCGTGTCAAATCTCCGMAAAGCTGGTCAGAAAACTCGCTATGAAA 1385
QY      966 CTCCTATGAAATCAAGAGCAGAGCTGTTCACCCATCGTGCAGCTACAGTACAAAGTTA 1025
Db      1384 TCACAAAGACTGATCCAAAGAGCTGAGCTGTCTACGCTCACTCCATTCAGTACATGTTA 1325
QY      1026 CGTAATGAGCATGGCTGATTAAGTTACAGGTGCGTTACATGGCAGCGTGTCAATTAAGGAG 1085
Db      1324 TGTCCGGGAACAGCTGCTGCTAACTCACTGCTGAGTTCAATGCAACGCTTCATTCGGGAG 1265
QY      1086 GCTGTGCTGTCTCAACGGTCTGG-----GAGCTACGGGAGGCTGTCACC--CCTGAG 1137
Db      1264 GCTGTCTGCTTTACGCACTGAGCACTACATAGAGAGCAAGTCTGTCACCTCCTTAAC 1205
QY      1138 CCCAGAGCTCGATCTCTTTAAGGACAAAGTCTCTCAACAGCTTAGTGTCTTAAGTGTTC 1197
Db      1204 TGCAGAGCTACCGTCTCTCTCAAGACGAAGTCTTTGCAAGTTCAGTGTCTCGTGTTC 1145
QY      1198 TCACGACAAAC---GCACTTAGTTTCAAGAGTATTTTGGCAATTTCTTAATCTGAGCAAGA 1254
Db      1144 TCGGCACAAACAATCGAGTGTAGTTTCAAGAGGATTTTGGCAACTCTTAACTGCAACAAGA 1085
QY      1255 ATAGGGG 1261
Db      1084 ATGGGG 1078

RESULT 11
AX408985/c
LOCUS      AX408985      1232 bp      DNA      linear      PAT 14-JUN-2002
DEFINITION Sequence 1632 from Patent WO0229103.
ACCESSION  AX408985
VERSION     AX408985.1  GI:21441690
KEYWORDS   Homo sapiens (human)
SOURCE      Homo sapiens
ORGANISM    Homo sapiens
REFERENCE   1
  AUTHORS   Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.
  TITLE     Gene expression profiles in liver cancer
  JOURNAL   Patent: WO 0229103-A 1632 11-APR-2002;
  GENE      GENE LOGIC INC (US)
FEATURES    Location/Qualifiers
             1..1232
             /organism="Homo sapiens"
             /mol_type="genomic DNA"
             /db_xref="taxon:9606"
             /note="EMBL/GenBank Accession No. D95274"
BASE COUNT 329 a 240 c 251 g 411 t 1 others
ORIGIN
Query Match      33.2%; Score 420.2; DB 6; Length 1232;
Best Local Similarity 67.5%; Pred. No. 9.7e-100;
Matches 848; Conservative 1; Mismatches 324; Indels 84; Gaps 15;

QY      53 ATTCAAGATTTTAAAGTCATACATGCAAAACATCTGCTAACTGCTGATTCGATTAAGAAAGAT 112
Db      1228 WTCCAGGATTTTAAAGTCATACATGCAAAACATATGCTAATTCATTCATTAAGAAAGAT 1169
QY      113 CAATGTAAACACTCCAAATTCGCACTGCAATTCGCAATGAAAAAGTTTCTTAGTGGT 172
Db      1168 CAATGTAAACACTCCAAATTCGCACTGCAATTCGCAATTTAAAAAATTCCTGTTAGTGGC 1109
QY      173 CGAAAGSCCCACACTGTGTTCTTGCCAGTGAGTTGTACAGAACGGCGTTAGCAC 232

```

```

Db      1108 TCAGGGTCCCACGCTGATTTCTCCAGTGAAGTTGTTACAGAAATCATGTCAGCAC 1049
QY      233 TAGGCG--TTGACAGAACTTCACAGACCCAAAGG-----TACCG 269
Db      1048 TAGACAGTTTACAGACCTTCACAGACCCAAAGAACATCAATAGGAAAGCAGCTACAG 989
QY      270 GAAGCATGTGTCGCGTGGGTGAGGTCTAGAGGGGGCGGCATCAATCATCATCAGAGTGT 329
Db      988 GAGGCGTGTGTCGCGTGGCGAGGTAAAGA--GGGTCAATTTGTCAGTACAGTGTCTC 930
QY      330 GGTACTCTGGCAAGACAGTGTG-----TTTCAGAAATATCTAAAAATAG 372
Db      929 GGTAACTCTGGCAAGACAGTGTATTAAGAAGGTTTCAAGAAATATCTAAAAATAT 870
QY      373 TTTAAAACTCTAAAGCGCAGCAGTGTATTTCTACACCCAGTGTACTTAGAAAAACGAAGG 432
Db      869 TTTAAAACTATTAAGCTGCAACACATGATTTTACACCTAGTTACTTAGAAAACTAAGGA 810
QY      433 AAGCACTAGTCAGCTGAGTAAAGGAAGGTGAAAAACAGGAACGCACTTCTACTATCTACCA 492
Db      809 AAGCACTATTAGTCTGAATA-----AAGTAACATGGAAGCACTTTTACTTAATCGACA 755
QY      493 AAAAAATCTCCGAATGCAATTAACAGAAAGATCTTATAGTACAGGTACAGATATGCTCG 552
Db      754 AAAAAACCTTCTAATGCAATTAACAGAAAGATTTTATAACAAG--GAGGCATATTGCTCA 696
QY      553 TTAAGAAGGGGCTCCTAAAGAAAAAGCACTTGTCTAAGTTAGCAACTGTGAGGATGGCAGT 612
Db      695 GTCAGAAGGGTCTTATAGAAAGCACTTACTAAGTTAGCGCTAACAGAAACCAACAGT 636
QY      613 TTAATAATGCACTCAACGCCCATCTGGGAGGAGCAGCAGGGGGAAGGGGGCTCAAGA 672
Db      635 TTAAGATGAATTAATGCCCAATTTGGGAGGCGATGGCAGGTGTAGAGAAAGAAAG 576
QY      673 GAGCACTGATPAGATCGGCCATTTGTCATCTACTGTTGACAGAAATTAACCGTTAAAA 732
Db      575 CTTAAGAAAAACATTTCTGATATAACCAACCTTTCTTTTCATCTACTGCTATTCAGAG 516
QY      733 AGCTTTTACCGTGACACTTTTATTCAGTTGAA---TTACTCCATGTACAAATGTTAGTAA 789
Db      515 AAATTAACCTTTTAGAGTTTTCACCGTGACACTTTCATTCCTTTGTACAAATGTTAGTGA- 457
QY      790 ATTAATCTCTACTTCTATA--TTAGTCAAAATACTGCTGTCTCTCTCTTTGATGACGTCGT 848
Db      456 ---AATCTCCATTCGTATTTTGCAAAATACTGCTCTTTGTCTCTTGTATCA----- 409
QY      849 TCACACTCTCCACCCAGCACACCCAGCTAGGAAACAGAAATATCTGTTAGAGGCAACAC 908
Db      408 -CACACACCCACCCCGGCACACCCACAGCTA--AACAGAAATTTCTTATTAGAGGAAATAG 352
QY      909 AGGAGCCAGAGTTCTGTTCAAAGCCTGCAGAGCCGGTGCAGCTGGTATTTTAGAGAACTC 968
Db      351 -----CAGTTCTGTTCAAATCTCCGAAAAGCTGCTGCAAGAACTCGCTATGAAATCA 298
QY      969 ACTATGAAATCAAGAGCAGAGCTGTTTACACCCATCGTGCAGTACAGTACAAAGTTACGT 1028
Db      297 CAAAGACTGATCCAAAGAGCTGAGCTGCTACGCTCACTCCATTACAGTACAAATGTTATGT 238
QY      1029 AATGAGCATGGGTGATTAAGTTACAGTGGTGTATACATGGCAGCGTGTCAATTAAGGAGGCT 1088
Db      237 CGGGAACACGTCGTGCTAACTCACTGCTGAGTTCAATGGCAACGCTTCATTCCGGAGGCT 178
QY      1089 GTCTGTGTACACAGGTCCTGG-----GAGCTACGGGAGGCTCGCACC--CCTGAGCCC 1140
Db      177 GTTCTGCTTTACCATCTGAGAACTACATAGGAGCAAGTGTCTGCACCTCTCTAAATGTC 118
QY      1141 AGAAGCTGCGAGTCTTCTTAAGGACAAAGTCTCTCAACAGCTTAGTGTCTACGTTGTTCTCA 1200
Db      117 AGAAGCTACGCTCTTCTCAAGACGAGGCTTTTGTAGTTCAGTGTCTCGGTTGTTCTCG 58
QY      1201 GCACAC---GCACCTAGTTCACAGGATTTTGGCAATTTTGGCAATTTCTTAATCTGAGCAAGA 1254

```

```

Db      57 GCACAACAATGTCAGTGTAGTTCAGAAAGGTATTTTGGCAACTCTTAACTCTGAACAAGA 1
RESULT 12
LOCUS   HUMPO2ST9/c
DEFINITION Homo sapiens mRNA, 1232 bp mRNA linear PRI 11-MAR-1998
ACCESSION D25274
VERSION   D25274.1 GI:464185
KEYWORDS
SOURCE    Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
          Ishida,Y., Hadano,S., Nagayama,T., Tomiyasu,H., Wakasa,K. and
          Ikeda,J.
          Isolation and characterization of 21 novel expressed DNA sequences
          from the distal region of human chromosome 4p
          Genomics 22 (2), 302-312 (1994)
          95104839
          7545969
          2 (bases 1 to 1232)
          Ishida,Y.
          Direct Submission
          Submitted (11-NOV-1993) Yoshikazu Ishida, Ikeda Genosphere
          project/ERATO/JRDC; Tokai University School of Medicine, Bohseidai,
          Isehara, Kanagawa 259-1193, Japan
          (E-mail:shinjienga.med.u-tokai.ac.jp, Tel:81-463-91-5095,
          Fax:81-463-91-4993)
          On Mar 25, 1994 this sequence version replaced gi:434743.
          Submitted (11-Nov-1993) to DDBJ by:
          Yoshikazu Ishida
          Ikeda Genosphere project/ERATO/JRDC
          Tokai University School of Medicine
          Bohseidai, Isehara
          Kanagawa 259-11
          Japan
          Phone: 0463-91-4056
          Fax: 0463-91-4110.
FEATURES
source      Location/Qualifiers
            1..1232
               /organism="Homo sapiens"
               /mol_type="mRNA"
               /db_xref="taxon:9606"
               /chromosome="4"
               /clone="PO2ST9"
               /tissue_type="brain striatum"
polyA_signal polyA signal
polyA_site    1214..1219
BASE_COUNT   329 a 240 c 251 g 411 t 1 others
ORIGIN
Query Match 33.2%; Score 420.2; DB 9; Length 1232;
Best local Similarity 67.5%; Pred. No. 9,7e-100;
Matches 848; Conservative 1; Mismatches 324; Indels 84; Gaps 15;
QY      53 ATTCAAGGATTTATTAGTCATACATGCAAAACATCTGCTAACTGCATTAGCAAAAGAT 112
Db      1228 WTCAGGATTTATTAGTCATACATGCAAAACATATGCTAATTGCAATTAGCAAAAGAT 1169
QY      113 CAATGTAAACAACTCCACAATTTCTGCACTGTCAATTGAAAAAGTTTCTTAGTGGT 172
Db      1168 CAATGTAAACAACTCCACAATTTCTGCACTGTCAATTGAAAAAGTTTCTTAGTGGC 1109
QY      173 CGAAGGCCCAACACTGTGTTCTTGCCAGTGTAGTGTGTCAGAACCGGCTTAGCAC 232
Db      1108 TGAAGGGTCCACGCTGTTTCTTCCGAGTGAGTTAAGTTGTACAGAACATCTGCAGCAC 1049
QY      233 TAGGCG--TTGACAGAACCTTCACAGACCCCAAGG-----TACCG 269
Db      1048 TAGCAGAGTTTACAGAACCTTCACAGACCCCAAGGACATCAATAGGCAAGGACTACAG 989

```

```

QY      270 GAAGCATGTGTCGGCTGGTGGTGTAGAGGGGGCGGCATCAATCAATGACAGTGT 329
Db      988 GAGGCGTGTGTCGGCTGGGCGAGGTAAAGA--GGGTCAGTATTGGTCAAGTGCAGTGC 930
QY      330 GGTACTCTGGGAAGACAGTGATG-----TTTCAGAAATATCTAAATATAG 372
Db      929 GGTAAATCTGGCAAGACAGTGATGTTAAGAAAGTTCATAGTTTAAAGAAATATCTCTAAATAT 870
QY      373 TTTAAAAAAGTAAAGCCGACAGTGTATTTCTACACCCAGTTACTAGAAAAACGAGGG 432
Db      869 TTTAAAAAAGTAAAGCTGCAACATGATTTTACACCTAGTTACTAGAAAACTAAGGA 810
QY      433 AAGCACTACTGACGTGAGTAAAGAGGTGAAACAGAGAACCGACTTCTACTATCTACCA 492
Db      809 AAGCACTTATTAGTCTGAATA-----AAGTAACATGGAAGACACTTTTACTAATCGACA 755
QY      493 AAAAATCTCCGATGCATATCAGAAAGATCTTATAGTACAGGTGAGCATATTTCTCTCG 552
Db      754 AAAAACCCTTCTAATGATTTATCAGAAAGATTTTATATACAAG--GAGGCAATATGCTCA 696
QY      553 TTAAGAAAGGGGTCTTAAAGAAAAAGCACTTGTCTAAGTTAGCAACTGTGAGATGGCCAGT 612
Db      695 GTCAAGAGGGGTCTTATAGAAAGACACTTACTAAGTTAGCGACTAACAGAACCAACAGT 636
QY      613 TTAATATGAGACTCAACGCCCATCTGGGAGGAGGACAGAGGGGGAGGAGGGGGCTCAAGA 672
Db      635 TTAAGATGAATTAATATGCCAATTTGGGAGGATGCGAGGTGTAAAGAAAAAGAAAAG 576
QY      673 GAGACACTGATAGATCGGCCATTTGTCTACTCTACTCTTTGACAGAAATTAACCGTTAAAA 732
Db      575 CTTAAGAAACATTTCTCTGATATATACCAACCTTTCTTTCATCATCTACTGCATTTGACAG 516
QY      733 AGCTTTACCCGTCGACACTTTTATTCAAGTGAA---TTACTCCATGTACAAATAGTGTAA 789
Db      515 AANTTAACCTTTTAGAGTTTACCCGTCGACACTTTCATCTCTGTACATGTAGTGA- 457
QY      790 ATTAAATCTTACTTCTAATA--TTAGTCAAAATACGTCTGCTCTCTTTGATGACGTCGTGTT 848
Db      456 ---AATCTCCACTTCGTATTTTGTCAAAATACTGTCTTTGTCTCTTTGATCA----- 409
QY      849 TCACACTCTCCACCCAGACACCCAGACTAGGAACAGATATCTCTGTAGAGGACACAC 908
Db      408 -CACACACCCACCCCGGACACCCACAGCTA--AACAGAAATCTTCTATTAGAGAAATAG 352
QY      909 AGGAGCCAGAGTTCTGTTCAAAGCCCTGCAGAACCCGTCAGCTGGTATTTTAGAGAACTC 968
Db      351 -----CAGTTCTGTTCAAAATCTCCGCAAAACSTGTCGAAAAACTCGCTATGAAATCA 298
QY      969 ACTATGAAATCAAGAGCAGAGCTGTTTACACCCATCGTGACGTACAGTACAAAGTTACGT 1028
Db      297 CAAAGACTGATCCAAAGAGCTGAGCTGCTACGCTCACTCCATTACAGTACAAATGTTATGT 238
QY      1029 AATGAGCATGGCTGATAGTTACAGGTGCTGTATACATGGCAGCGTGTCTATTAGGAGGCT 1088
Db      237 CGGGAACAGCTGTCTTAATCTCACTGTGAGTTCAATGGCAACGTTTCATTCGGAGGCT 178
QY      1089 GTCTGTGTGCACACGCTCTGG-----GAGCTACGGGAGGGTCTGCACC--CTGAGCCCC 1140
Db      177 GTTCTGTTTACGATCTGAGAACTACATAGGAGAGCAAGTGTCTGCACCTCTCTAATCTGC 118
QY      1141 AGAAGCTGAGCTTCTTAAAGGACAAAGTCTTCAACAGCTTAGTCTTACGTTGTTCTCA 1200
Db      117 AGAAGTACGCTCTTCTCAAGACGAGGTCTTTTGTAGTTCAAGTCTCGGTCTCGTCTCG 58
QY      1201 GCACAAAC---GCAACTTAGTTTCAAGGATTTTGGCAATTTCTTAATCTGAGCAAGA 1254
Db      57 GCACAAACATGAGTGTAGTTTCAAGAGTATTTTGGCAACTCTTAACTCTGACAAAGA 1

```

RESULT 13

AC102775/c

LOCUS

DEFINITION

Mus musculus clone RP23-115C10, WORKING DRAFT SEQUENCE, 9 unordered

AC102775 185245 bp DNA linear HTG 23-MAR-2003

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

pieces.
AC102775.3 GI:29164654
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
Mus musculus (house mouse)
Mus musculus

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 185245)
Birren, B., Nusbaum, C. and Lander, E.
Mus musculus, clone RP23-115C10
Unpublished
2 (bases 1 to 185245)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Chapel, Y., Colangelo, M., Collins, S., Collumore, A., Cook, A.,
Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K.,
Lanazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
Maclean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPeeters, R., Meldrim, J.,
Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zemek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
REFERENCE
AUTHORS

Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 185245)

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,
Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,
Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y.,
Collumore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K.,
Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,
Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S.,
Graham, L., Grand-Pierre, N., Hagel, N., Hagopian, D., Hagos, B.,
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., MacLean, C.,
Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J.,
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C.,
Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,
Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,
Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zemek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (23-MAR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 23, 2003 this sequence version replaced gi:22381790.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center

TITLE
JOURNAL
COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L19647

Center clone name: 115_C10
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 183758 bases at least Q40
Consensus quality: 184136 bases at least Q30
Consensus quality: 184291 bases at least Q20
Insert size: 178000; agarose-fp
Quality coverage: 12.6 in Q20 bases; agarose-fp
Quality coverage: 12.2 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 53693: contig of 53693 bp in length
53694 53793: gap of 100 bp
53794 56675: contig of 2882 bp in length
56676 56775: gap of 100 bp
56776 60589: contig of 3814 bp in length
60590 60689: gap of 100 bp
60690 67210: contig of 6521 bp in length
67211 67310: gap of 100 bp
67311 82585: contig of 15275 bp in length
82586 82685: gap of 100 bp
82686 103118: contig of 20433 bp in length
103119 103218: gap of 100 bp
103219 132145: contig of 28927 bp in length
132146 132245: gap of 100 bp
132246 182335: contig of 50090 bp in length
182336 182435: gap of 100 bp
182436 185245: contig of 2810 bp in length.

FEATURES
source
1. 185245
/organism="Mus musculus"
/db_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP23-115C10"
/clone_lib="RPCI-23 Female Mouse BAC"
1. 53693
/note="assembly_fragment"
clone_end:SP6
vector_side:left
53794..56675
/note="assembly_fragment"
56776..60589
/note="assembly_fragment"
60590..67210
/note="assembly_fragment"
67311..82585
/note="assembly_fragment"
82686..103118
/note="assembly_fragment"
103219..132145
/note="assembly_fragment"
132246..182335
/note="assembly_fragment"
182436..185245
/note="assembly_fragment"
clone_end:T7
vector_side:right

BASE COUNT 56870 a 34956 c 35069 g 57550 t 800 others
ORIGIN
Query Match 32.2%; Score 408.2; DB 2; Length 185245;
Best Local Similarity 89.8%; Pred. No. 2.2e-96;
Matches 486; Conservative 0; Mismatches 43; Indels 12; Gaps 4;

ORIGIN

Query Match 29.2%; Score 369.4; DB 2; Length 192498;
Best Local Similarity 83.6%; Pred. No. 3.9e-86;
Matches 494; Conservative 0; Mismatches 56; Indels 41; Gaps 5;

QY 35 TTTTTCCTTTTTCCTTATGATTCAGGATTTATTAGTCATACATGCAAAACATCTACTGCTA 94
DB 90240 TTTTCCTTTTTCCTTATGATTCAGGATTTATTAGTCATACATGCAAAACATCTACTGCTA 90299

QY 95 ACTGCATTAGCAAAAGATCAATGTAAACACATCCCAATTTCTGCAACTGTCAATGAAA 154
DB 90300 ACTGCATTAGCAAAAGATCAATGTAAACACATCCCAATTTCTGCAACTGTCAACTGAAA 90359

QY 155 AAGTTTGTTCCTAGTGTGCGAAGGCCCAACACTGTGTTCTTCCAGTGTAGTTAGTTGT 214
DB 90360 AAGTTTGTTCCTAGTGTGCGAAGGCCCAACACTGTGTTCTTCCAGTGTAGTTAGTTGT 90419

QY 215 ACAGAACGGCGTTAGCACTAGCGCTTTGACAGAACCTTCACAGACCCCAAGG----- 264
DB 90420 ACAGAACGGCGTTAGCACTAGCGCTTTGACAGAACCTTCACAGACCCCAAGGACATCTCTA 90479

QY 265 -----TACCGGAAGCATGTGTCCGCGTGGGTGAGTCTAGAGGGGGCGGCATCA 313
DB 90480 GGCAGAGCCACGACAGGAAGCGTCTCTCTCGTGTGAGTGTAGTCTAGA-GAGGCGAGCATTA 90538

QY 314 ATCACATGACAGTGTGTTGTTACTCTGGCAAGACAGTGTGTTTCAGAAATATCTAAAAATAGT 373
DB 90539 GTCCATGACAGTGTGTTGTTACTCTGGCAAGACAGTGTGTTTCAGAAATATCTAAAAATAGT 90589

QY 374 TTAAAAACTGTAAAGCCGACGACGTGATTTCTTACACCCAGTGTCTAGAAAAACGAAGGGA 433
DB 90590 TTAAAAACTGTAAAGCCGACGACGTGATTTCTTACACCCAGTGTCTAGAAAAACGAAGGGA 90649

QY 434 AGCACTAGTCAGC--TGAGTAAAGGAGGTGAAACAGGACCGACCTTCTACT----- 484
DB 90650 AGCACTAGTCAGC--TGAGTAAAGGAGGTGAAACAGGACCGACCTTCTACTGATCGAC 90709

QY 485 -ATCTACCAAAAAATCTCCGAATGCATTATCAGAAAGATCTTATAGTACAGGTCTAGACA 543
DB 90710 CAAAAAATAAAAAATCTCTGAATGCATTATCAGAAAGATCTTATAGTACAGGTCTAGACA 90769

QY 544 TATTGCTCGTTAAGAGGGGGTCTTAAAGAAAAGCACTTGCTAAGTTAGCA 594
DB 90770 TATTGCTCATTAAGAGGGGGTCTTAAAGAAAAGCACTTGCTAAGTTAGCA 90820

Search completed: November 23, 2003, 14:01:54
Job time : 4753.05 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 23, 2003, 09:17:01 ; Search time 289.886 Seconds
(without alignments)
9470.385 Million cell updates/sec

Title: US-09-717-321A-15

Perfect score: 1017

Sequence: 1 cccctattctgtcagatt.....ccttgggtctgtgaggttc 1017

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_19Jun03.*
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1017	100.0	1017	22	AAH22397
2	1017	100.0	1266	22	AAH22396
3	290.6	28.6	1022	22	AAH22398
4	290.6	28.6	2051	21	AAF18236
5	290.6	28.6	28567	25	ABT17030
6	285.2	28.0	1232	24	ABK83623
7	285.2	28.0	1232	24	ABN95134
8	193.2	19.0	605	24	ABL38065

C	9	187.2	18.4	605	21	AAA16478
10	172.2	16.9	432	25	ABX42447	
11	168	16.5	1318	23	AA865569	
12	164	16.1	2167	24	ABZ35389	
13	126.4	12.4	285	24	ABL71132	
C	14	123.2	12.1	416	25	ABX43616
15	113	11.1	348	22	AAH22399	
16	111.6	11.0	447	21	AAA89693	
17	111.2	10.9	422	21	AAA89694	
18	103.2	10.1	3740	22	AAH07296	
19	98.2	9.7	581	24	ABV87661	
C	20	91.6	9.0	565	24	ABV86985
C	21	87.8	8.6	506	22	AAH08004
C	22	87.6	8.6	470	22	AAH18145
C	23	85.6	8.4	466	22	AAH25750
C	24	85.6	8.4	495	24	ABL37914
C	25	68	6.7	588	21	AAA16128
26	67	6.6	594	24	ABN60324	
27	62.2	6.1	426	25	ABX42177	
C	28	61.4	6.0	378	21	AAC06949
C	29	59	5.8	404	25	ABX49041
C	30	52.4	5.2	162	24	ABK45998
31	41.4	4.1	901	25	ABT16468	
C	32	39.2	3.9	4081	23	ABL12618
C	33	37.4	3.7	13712	24	ABJ33531
34	37	3.6	352	22	AAH22400	
35	37	3.6	2277	25	ABZ58242	
C	36	37	3.6	260209	24	AB856564
C	37	36.8	3.6	3014	23	ABL06602
C	38	36.8	3.6	3724	23	ABL20520
C	39	36.6	3.6	995	22	AAH15320
40	36.4	3.6	7892	24	ABK40055	
C	41	36	3.5	5098	22	AAH13425
C	42	36	3.5	14041	22	AAH48024
C	43	36	3.5	16258	24	ABL70376
C	44	36	3.5	16258	24	ABK40038
C	45	36	3.5	611590	21	AAF22303

ALIGNMENTS

RESULT 1
AAH22397
ID AAH22397 standard; DNA; 1017 BP.
XX AC
XX AAH22397;
XX AC
XX DT 22-AUG-2001 (first entry)
XX DE
XX DE Human rac1 gene related nucleotide sequence #1.
XX Identification; toxic; hepatotoxic; differential gene expression;
KW NSAIID; non-steroidal antiinflammatory drug; ds.
XX OS
XX Homo sapiens.
XX PN
XX WO200138579-A2.
XX PD
XX 31-MAY-2001.
XX PF
XX 21-NOV-2000; 2000WO-US32049.
XX PR
XX 22-NOV-1999; 99US-0166923.
XX PR 18-FEB-2000; 2000US-0183531.
XX PR 20-NOV-2000; 2000US-0717321.
XX {CURA-} CURAGEN CORP.
XX PI
XX Gould-Rothberg BE, Dipippo VA, Ramseil TM, Gerwein RW;
XX WPI; 2001-355948/37.

Human colon cancer
Bovine EST associa
DNA encoding novel
Human gene express
Corn tassal-derive
Bovine EST associa
Human rac1 gene re
Mouse Exo103 nucle
Mouse Rab2 nucleot
Human reproductive
Human colon cancer
Human colon cancer
Human breast cance
Human breast cance
Human breast cance
Human colon tumor
Human colon cancer
Human cancer relat
Bovine EST associa
Human secreted pro
Bovine EST associa
cDNA encoding colo
Human intracellular
Drosophila melanog
Human immune syste
Human rac1 gene re
Human G-protein co
Human SULF2 genom
Drosophila melanog
Drosophila melanog
Human breast cance
Human chemically p
Murine Kit/stem ce
Internal control B
Chemically treated
Human chemically p
Arabidopsis thalia

PT Screening hepatotoxic agent comprises contacting test cell population
PT expressing RISKMARKER or INJURYMARKER with agent, comparing expression
PT with reference population and identifying difference in expression
PT levels -
XX
XX Disclosure; Page 7-8; 76pp; English.
PS
XX The present invention describes a method of screening a test agent for
CC hepatotoxicity. The method comprises: (a) providing a test cell
CC population comprising a cell capable of expressing one or more nucleic
CC acid sequences selected from the group consisting of RISKMARKER 1-8
CC and INJURYMARKER 1-10; (b) contacting the test cell population with a
CC test agent; (c) measuring expression of one or more of the nucleic
CC acid sequences in the test cell population; (d) comparing the
CC expression of the nucleic acid sequence in the test cell population to
CC the expression of the nucleic acid sequence in an reference cell
CC population comprising at least one cell whose exposure status to a
CC hepatotoxic agent is known; and (e) identifying a difference in
CC expression levels of the RISKMARKER or INJURYMARKER sequences, if
CC present, in the test cell population and reference cell population.
CC The method is useful for identifying a hepatotoxic agent. The present
CC sequence is given in the exemplification of the present invention.
XX
XX Sequence 1017 BP; 245 A; 245 C; 216 G; 311 T; 0 other;
SQ

Query Match 100.0%; Score 1017; DB 22; Length 1017;
Best Local Similarity 100.0%; Pred. No. 6.9e-308;
Matches 1017; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCCCTATTCTCTCAGATTAAGATTGCGAAATACCTTGTGAACATGTTGCGTTGTG 60
DB 1 CCCCTATTCTCTCAGATTAAGATTGCGAAATACCTTGTGAACATGTTGCGTTGTG 60
QY 61 CTGAGAACAGTAAGACATTAAGCTGTGTGAGAGACTTTGTCTTAAAGAGACTGCACCTTC 120
DB 61 CTGAGAACAGTAAGACATTAAGCTGTGTGAGAGACTTTGTCTTAAAGAGACTGCACCTTC 120
QY 121 TGGGCTCAGGGTGCAGACCCCTCCCGTAGTCCAGACCGTGTGACACGACAGCCTCC 180
DB 121 TGGGCTCAGGGTGCAGACCCCTCCCGTAGTCCAGACCGTGTGACACGACAGCCTCC 180
QY 181 TTAATGACACGCTGCGATGTAAGCAGCTGTAATATCAGCCCATGCTCATTACCTAAC 240
DB 181 TTAATGACACGCTGCGATGTAAGCAGCTGTAATATCAGCCCATGCTCATTACCTAAC 240
QY 241 TTGTGACTGTACGTACAGATGGGTGTAACAGCTCTCTCTTGAATTCATAGTGAATCT 300
DB 241 TTGTGACTGTACGTACAGATGGGTGTAACAGCTCTCTCTTGAATTCATAGTGAATCT 300
QY 301 CTAAATACAGCTGACCGGCTTCTCAGGCTTTGAACAGAACTCTGGCTCCTGTGTGC 360
DB 301 CTAAATACAGCTGACCGGCTTCTCAGGCTTTGAACAGAACTCTGGCTCCTGTGTGC 360
QY 361 CTCTACGAGATATTCTGTCTAGTCTGGGTGTGCTGGGTGAGTGTGTAACACAGCA 420
DB 361 CTCTACGAGATATTCTGTCTAGTCTGGGTGTGCTGGGTGAGTGTGTAACACAGCA 420
QY 421 CGTTCATCAAGGAGACAGACAGTATTGTGACTAATATGAAGTAGAGATTAAATTACACTA 480
DB 421 CGTTCATCAAGGAGACAGACAGTATTGTGACTAATATGAAGTAGAGATTAAATTACACTA 480
QY 481 CATGTACATGGAGTAATCAACTGAATAAAGTGCACGGTAAGACTTTTAAACGGTT 540
DB 481 CATGTACATGGAGTAATCAACTGAATAAAGTGCACGGTAAGACTTTTAAACGGTT 540
QY 541 AATTTCTGTCAAAAGTAGATGACAAATGGCCCATCTTATCAGTGTCTCTCTTGAGCCCC 600
DB 541 AATTTCTGTCAAAAGTAGATGACAAATGGCCCATCTTATCAGTGTCTCTCTTGAGCCCC 600
QY 601 CTTTCCCTCTGCTGCTCCCTCCAGATGGGGCGTGTAGTCCATATTAACTGCGCATCC 660
DB 601 CTTTCCCTCTGCTGCTCCCTCCAGATGGGGCGTGTAGTCCATATTAACTGCGCATCC 660

QY 661 TCACAGTTGCTAACTTAGCAAGTGCTTTTCTTTAGGACCCCTCTTAAACGAGCAATATG 720
DB 661 TCACAGTTGCTAACTTAGCAAGTGCTTTTCTTTAGGACCCCTCTTAAACGAGCAATATG 720
QY 721 TCTGACCTGTACTATAAAGATCTTTCTGATAATGCAATCGGAGATTTTGTGTAGATAGT 780
DB 721 TCTGACCTGTACTATAAAGATCTTTCTGATAATGCAATCGGAGATTTTGTGTAGATAGT 780
QY 781 AGAAGTGGTTCCTGCTTTTACCTTCTTACTCAGCTGACTAGTGTCCCTTCGTTTT 840
DB 781 AGAAGTGGTTCCTGCTTTTACCTTCTTACTCAGCTGACTAGTGTCCCTTCGTTTT 840
QY 841 CTAGTAACCTGGGTGTAGAAATCACGCTGCTGGGCTTTACAGTTTAAACTATTTAGAT 900
DB 841 CTAGTAACCTGGGTGTAGAAATCACGCTGCTGGGCTTTACAGTTTAAACTATTTAGAT 900
QY 901 ATTCTGAAACATCACTGCTTGTCCAGAGTACCAACACTGTATGATGATGCGGCCCC 960
DB 901 ATTCTGAAACATCACTGCTTGTCCAGAGTACCAACACTGTATGATGATGCGGCCCC 960
QY 961 CTCTAGACCTCAACCCAGCGGACATGCTTCCGGTACCTTTGGGTCTGTGAGGTTTC 1017
DB 961 CTCTAGACCTCAACCCAGCGGACATGCTTCCGGTACCTTTGGGTCTGTGAGGTTTC 1017

RESULT 2

AAH22396/c
ID AAH22396 standard; DNA; 1266 BP.

XX AAH22396;

DT 22-AUG-2001 (first entry)

XX Human rac1 contig SEQ ID NO:2.

XX Identification; toxic; hepatotoxic; differential gene expression;
KW NSAID; non-steroidal antiinflammatory drug; ds.

XX Homo sapiens.

XX WO200138579-A2.

XX 31-MAY-2001.

XX 21-NOV-2000; 2000WO-US32049.

XX 22-NOV-1999; 99US-0166923.

XX 18-FEB-2000; 2000US-0183531.

XX 20-NOV-2000; 2000US-0717321.

XX (CURA-) CURAGEN CORP.

XX Gould-Rothberg BE, Dipippo VA, Ramseh TM, Gerwein RW;

XX WPI; 2001-355948/37.

XX Screening hepatotoxic agent comprises contacting test cell population
PT expressing RISKMARKER or INJURYMARKER with agent, comparing expression
PT with reference population and identifying difference in expression
PT levels -

XX Disclosure; Page 7; 76pp; English.

XX The present invention describes a method of screening a test agent for
CC hepatotoxicity. The method comprises: (a) providing a test cell
CC population comprising a cell capable of expressing one or more nucleic
CC acid sequences selected from the group consisting of RISKMARKER 1-8
CC and INJURYMARKER 1-10; (b) contacting the test cell population with a
CC test agent; (c) measuring expression of one or more of the nucleic
CC acid sequences in the test cell population; (d) comparing the
CC expression of the nucleic acid sequence in the test cell population to
CC the expression of the nucleic acid sequence in an reference cell
CC population comprising at least one cell whose exposure status to a

CC hepatotoxic agent in known; and (e) identifying a difference in
 CC expression levels of the RISKMARKER or INJURYMARKER sequences, if
 CC present, in the test cell population and reference cell population.
 CC The method is useful for identifying a hepatotoxic agent. The present
 CC sequence is given in the exemplification of the present invention.

XX
 SQ Sequence 1266 BP; 385 A; 258 C; 285 G; 338 T; 0 other;
 Query Match 100.0%; Score 1017; DB 22; Length 1266;
 Best Local Similarity 100.0%; Pred. No. 7.7e-308;
 Matches 1017; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CCCCTATTCTTGGCTCAGATTAAAGATTGCGAAATGCGAAATACCTTGTGAACTAAGTTGGTCTG 60
 DB 1261 CCCCTATTCTTGGCTCAGATTAAAGATTGCGAAATACCTTGTGAACTAAGTTGGTCTG 1202
 QY 61 CTGAGAACACGTAAGCACTAAGCTGTGTTGAGAGACTTGTCTTAAAGAACTGCGAGTTTC 120
 DB 1201 CTGAGAACACGTAAGCACTAAGCTGTGTTGAGAGACTTGTCTTAAAGAACTGCGAGTTTC 1142
 QY 121 TGGGCTCAGGGGTGCGAGACCTCCCGTAGCTCCGAGACCGGTGTCACACAGCAGACCTCC 180
 DB 1141 TGGGCTCAGGGGTGCGAGACCTCCCGTAGCTCCGAGACCGGTGTCACACAGCAGACCTCC 1082
 QY 181 TTAATGACAGCTGCCATGTAAACGACCTGTAACTTATCAGCCCATGCTCATACGTAAC 240
 DB 1081 TTAATGACAGCTGCCATGTAAACGACCTGTAACTTATCAGCCCATGCTCATACGTAAC 1022
 QY 241 TTTGTACTGTACGTACAGTGGGTGTAAACAGCTCTGCTCTTTGATTTCATAGTGAGTTCT 300
 DB 1021 TTTGTACTGTACGTACAGTGGGTGTAAACAGCTCTGCTCTTTGATTTCATAGTGAGTTCT 962
 QY 301 CTAATAATACAGCTGACCGGCTTCTGAGGCTTGTAAACAGCACTCTGGCTCTGTGTTC 360
 DB 961 CTAATAATACAGCTGACCGGCTTCTGAGGCTTGTAAACAGCACTCTGGCTCTGTGTTC 902
 QY 361 CTCTAAGCAAGTATTCCTGTTCTTCTAGTCTGGGTGTGCTGGGTGAGTGTGTGAACACGA 420
 DB 901 CTCTAAGCAAGTATTCCTGTTCTTCTAGTCTGGGTGTGCTGGGTGAGTGTGTGAACACGA 842
 QY 421 CGTCAATCAAGGAGACAGACAGTATTTGTAATAATGAAGTAGAGATTAAATTACACTA 480
 DB 841 CGTCAATCAAGGAGACAGACAGTATTTGTAATAATGAAGTAGAGATTAAATTACACTA 782
 QY 481 CATTTGACATGGAGTAATTCACGTGATTAAGTGTACAGTGTCTCTCTTTTAAACGGTT 540
 DB 781 CATTTGACATGGAGTAATTCACGTGATTAAGTGTACAGTGTCTCTCTTTTAAACGGTT 722
 QY 541 AATTTCTGTCAACAGTAGATGACAAATGGCCGATCTTATCAGTGTCTCTCTTTGAGCCCC 600
 DB 721 AATTTCTGTCAACAGTAGATGACAAATGGCCGATCTTATCAGTGTCTCTTTGAGCCCC 662
 QY 601 CCTTCCCGCTGTCTCCCTCCAGATGGGGCTTGAGTCCATATTTAACTGGCCATCC 660
 DB 661 CCTTCCCGCTGTCTCCCTCCAGATGGGGCTTGAGTCCATATTTAACTGGCCATCC 602
 QY 661 TCACAGTTGCTAACTTAGCAAGTCTTTCTTTAGGACCCCTCTCTTAAGCAATATG 720
 DB 601 TCACAGTTGCTAACTTAGCAAGTCTTTCTTTAGGACCCCTCTCTTAAGCAATATG 542
 QY 721 TGTGACCTGTACTATAAGATCTTTCTGATAATGCAATTCGAGATTTTGTGTAGATAGT 780
 DB 541 TGTGACCTGTACTATAAGATCTTTCTGATAATGCAATTCGAGATTTTGTGTAGATAGT 482
 QY 781 AGAAGTGCCTTCTGTTTTCACCTTCTTACTCAGTGTAGTGTCTTCCCTTCTGTTT 840
 DB 481 AGAAGTGCCTTCTGTTTTCACCTTCTTACTCAGTGTAGTGTCTTCCCTTCTGTTT 422
 QY 841 CTAGTAACCTGGGTGTAGAAATCAGGTGCTGCGGCTTTACAGTTTAACTATTTTAGAT 900
 DB 421 CTAGTAACCTGGGTGTAGAAATCAGGTGCTGCGGCTTTACAGTTTAACTATTTTAGAT 362
 QY 901 ATTCTGAACATCACTGTCTTGGCCAGAGTACCACCACTGTCAITGTGATTCGCCGCC 960

DB 361 ATTTGAAACATCACTGTCTTCCAGAGTACCAACACTGTATGTATGATGCGGCCCC 302
 QY 961 CTCTAGACCTACCCACCGGACACATCTTCCGTAACCTTTGGCTGTGTGAGTTTC 1017
 DB 301 CTCTAGACCTACCCACCGGACACATCTTCCGTAACCTTTGGCTGTGTGAGTTTC 245
 RESULT 3
 AAH22398
 ID AAH22398 standard; DNA; 1022 BP.
 XX AAH22398;
 XX 22-AUG-2001 (first entry)
 XX Human rac1 gene related nucleotide sequence #2.
 XX Identification; toxic; hepatotoxic; differential gene expression;
 XX NSAID; non-steroidal antiinflammatory drug; ds.
 XX Homo sapiens.
 XX WO200138579-A2.
 XX 31-MAY-2001.
 XX 21-NOV-2000; 2000WO-US32049.
 XX 22-NOV-1999; 99US-0166923.
 XX 18-FEB-2000; 2000US-0183531.
 XX 20-NOV-2000; 2000US-0717321.
 XX (CURA-) CURAGEN CORP.
 XX Gould-Rothberg BE, Dipippo VA, Ramseh TM, Gerwein RW;
 XX WPI; 2001-355948/37.

Screening hepatotoxic agent comprises contacting test cell population
 expressing RISKMARKER or INJURYMARKER with agent, comparing expression
 with reference population and identifying difference in expression
 levels -

Disclosure; Page 7-8; 76pp; English.

The present invention describes a method of screening a test agent for
 hepatotoxicity. The method comprises: (a) providing a test cell
 population comprising a cell capable of expressing one or more nucleic
 acid sequences selected from the group consisting of RISKMARKER 1-8
 and INJURYMARKER 1-10; (b) contacting the test cell population with a
 test agent; (c) measuring expression of one or more of the nucleic
 acid sequences in the test cell population; (d) comparing the
 expression of the nucleic acid sequence in the test cell population to
 the expression of the nucleic acid sequence in a reference cell
 population comprising at least one cell whose exposure status to a
 hepatotoxic agent is known; and (e) identifying a difference in
 expression levels of the RISKMARKER or INJURYMARKER sequences, if
 present, in the test cell population and reference cell population.
 The method is useful for identifying a hepatotoxic agent. The present
 sequence is given in the exemplification of the present invention.

Sequence 1022 BP; 273 A; 208 C; 205 G; 336 T; 0 other;

Query Match 28.6%; Score 290.6; DB 22; Length 1022;
 Best Local Similarity 64.5%; Pred. No. 3.1e-80;
 Matches 664; Conservative 0; Mismatches 304; Indels 61; Gaps 13;

QY 1 CCCCTATTCTTGGCTCAGATTAAAGATTGCGAAATACCTTGTGAACTAAGTTGC---GTT 57
 DB 1 CCCCTATTCTTGGCTCAGATTAAAGATTGCGAAATACCTTGTGAACTAAGTTGTT 60
 QY 58 GTGCTGAGAACACGTAAGCACTTAAGTCTGTGAGAGACTTTGTCTTAAAGAACTGCAGC 117

Db 935 GAAAGAGCTCCCGAATGAAGGTTGCCATTGAATCACCAGTGAAGTACGACGAGTGT 994
QY 230 CATTACGTAACCTTTGACTGTAGCTCAGTCAGATGGGTGTAACAGCTCTGCTCTTTGATTCA 289
Db 995 CCGACATACCAATGTAATGTAATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1054
QY 290 TAGTGAGTCTCTAAATACCAATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 349
Db 1055 TTTGTGATTTATAGCAGGTTTCTGACCACTTTGCGGAGATTTGAACAGAACTG-- 1112
QY 350 TCTGTGTGCTCTAAGCAAGTATTTCTGTTCTAGTGTGGTGTGGTGTGGTGTGGTGT 409
Db 1113 ---CTATTTCTCTAATGAAGATTTCTGTT--TAGCTGTGGTGTGGTGTGGTGTGGTGT 1166
QY 410 GTGAACACAGCTCATCAAGGAGACAGACAGTATTTTAC--TAATATGAAGTAGAGAT 468
Db 1167 GT-----GTGATCAAGGACAAAGACAGTATTTTGAACAAATACGAAGTGGAG-- 1214
QY 469 TAATTTACACTACATTTGATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 525
Db 1215 --ATTACACTACATTTGATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1272
QY 526 AGCTTTTAAAGTGAATTTCTGTTCTGTTCAACAGTGAAGTGAAGTGAAGTGAAGTGAAGT 585
Db 1273 ATTTCTGCAATGCACTAGATGATGAAGAAAGTGTGTTATATCAGAAATGTTTCT 1332
QY 586 TCTCTCTGAGCCCTTCCCTCTGCTGCTCCCTCCAGTGGGGTGTGAGTGCATAT 645
Db 1333 TAAGCTTTTCTTCTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1392
QY 646 TTAACCTGGCCATCCTCACAGTTGCTAACTTAGCAAGTGTCTTTCTTTAGGACCCCTTC 705
Db 1393 TTAACCTGGTGTCTGTTAGTCTGCTAACTTAGCAAGTGTCTTTCTTTAGGACCCCTTC 1452
QY 706 TTAACGAGCAATGCTGACCTGTACTATAGATCTTTCTGATAATGCAATTCGGAGATT 765
Db 1453 TGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1511
QY 766 TTTTGTGATAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 825
Db 1512 TTTTGTGATAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1566
QY 826 GCTTCTCTCTGTTTCTAGTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 885
Db 1567 GCTTCTCTGTTTCTAGTAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1626
QY 886 TAAACTATTTTGAATA-----TTCTGAACATCACTGTCTTGGCCAGAG 928
Db 1627 TAAATATTTTGAATAATTTTAAATATGAACTTCTTAACTATCACTGTCTTGGCCAGAT 1686
QY 929 TACCAACACTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 988
Db 1687 TACCAACACTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1745
QY 989 CTTCCGGTA 997
Db 1746 CTTCTGTA 1754

RESULT 5
ABT17030
ID ABT17030 standard; DNA; 28567 BP.
XX
AC ABT17030;
XX
DT 03-APR-2003 (first entry)
XX
DE Human MP21 gene Rac1 SEQ ID No 4.
XX
KW Cytostatic; p21 pathway modulating agent; cancer; angiogenic; apoptotic;
XX cell proliferation disorder; MP21; gene; ds.

OS Homo sapiens.
XX WO2003006990-A1.
XX 23-JAN-2003.
XX 10-JUL-2002; 2002WO-US21549.
XX 12-JUL-2001; 2001US-305017P.
XX 10-OCT-2001; 2001US-328491P.
XX 15-FEB-2002; 2002US-357452P.
XX (EXEL-) EXELIXIS INC.
XX Friedman L, Plowman GD, Belvin M, Li D, Funke RP;
XX WPI; 2003-221779/21.
XX P-PSDB; ABJ19756.
XX
XX Identifying candidate p21 pathway modulator, by contacting an assay
XX system having modifiers of p21 polypeptide or gene with a test agent to
XX provide a reference activity in system and detecting test agent-biased
XX activity -
XX Examples; Page 56-72; 199pp; English.
XX
XX The invention relates to a novel method for identifying a candidate p21
XX pathway modulating agent. The novel method comprises contacting an assay
XX system, comprising a purified MP21 polypeptide (modifier of p21) or
XX nucleic acid, with a test agent under conditions, so that but for the
XX presence of a test agent, the assay system provides a reference activity
XX and detection of test agent-biased activity of the assay system. The
XX novel method of the invention is useful for identifying a candidate p21
XX pathway modulating agent. The invention also includes a method for
XX modulating the p21 pathway of a cell, and a method for diagnosing a
XX disease e.g. cancer in a patient. The identified modulators are useful in
XX diagnosis, therapy and pharmaceutical development. The modulators are
XX useful in a variety of diagnostic and therapeutic applications including
XX angiogenic, apoptotic and cell proliferation disorders. This
XX polynucleotide sequence represents a gene encoding an MP21 protein of the
XX invention.
XX
XX Sequence 28567 BP; 6762 A; 6358 C; 6796 G; 8651 T; 0 other;
XX
XX Query Match 28.6%; Score 290.6; DB 25; Length 28567;
XX Best Local Similarity 64.5%; Pred. No. 2e-79;
XX Matches 664; Conservative 0; Mismatches 304; Indels 61; Gaps 13;
XX
QY 1 CCCCTATCTTCTGCTCAGATTAGAATTGCCAAATACCTTGTGAACCTACACTGCATTGT 57
Db 26871 CCCCCATCTTCTGCTCAGATTAGAATTGCCAAATACCTTGTGAACCTACACTGCATTGT 26930
QY 58 GTGCTGAGAACACGTAAGCACTAAGCTGTGTGAGACTTTGCTTAAAGAACTCGAGC 117
Db 26931 GTGCCGAGAACCCGAGCACTGAACCTTTGCAAGAACCTTCGTTTGAAGACGCTAGC 26990
QY 118 TTCTGGGCTCAGG-GGTGCAGACCCCTCCCGTAGC-----TCCCAGACCGGTGACACA 169
Db 26991 TTCTGCAGTTAGGAGGTGCAGACACTTGTCTCTCTATGTAGTTCTCAGATGCGTAAGCA 27050
QY 170 GCACAGCCTCTTAATGACACGCTGCCATGTAAACGACCTGTAACTTATCAGCCCATGCT 229
Db 27051 GAACAGCCTCCGGAATGAAGCGTTGCCATTGAACCTACCAGTGAAGTACACACGCTGT 27110
QY 230 CATTACGTAACCTTTGACTGTAGCTCAGATGAGTGAAGTGAAGTGAAGTGAAGTGAAGT 289
Db 27111 CCGGACATAAACAATTGACTGTCTAATGAGTGAAGTGAAGTGAAGTGAAGTGAAGT 27170
QY 290 TAGTGAGTCTCTAAATACCAATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 349
Db 27171 TTTGTGATTTATAGCAGGTTTCTGACCACTTTTGGGAGATTTTGAACAGAACTG-- 27228
QY 350 TCTGTGTGCTCTAAGCAAGTATTTCTGTTCTAGTGTGGTGTGGTGTGGTGTGGTGTGGTGT 409

Db 27229 -----CTATTTCTCTTAATGAAGAAATTCGTT--TAGCTGTGGTGTGCGGGTGGGTGT 27282
Qy 410 GTGAACACAGCGTCATCAAAAGGAGACAGACAGTATTTTGAC-TAATATGAAGTAGAGAT 468
Db 27283 GT-----GTCATCAAGGACAAAGACAGTATTTTGACAAATACGAAGTGGAG-- 27330
Qy 469 TAATTTACATACATTTGACATGAGTAA-----TTCACTGAATAAAGTGTCCAGGTAA 525
Db 27331 --ATTACATACATTTGACAGGAATGAAGTGTACGGGTAAAACTCTAAAGGTTA 27388
Qy 526 AGCTTTTAAACGGTTAATTTCTGTCAACAGTAGATGACAAATGCCCATCTTATCACTG 585
Db 27389 ATTCTGTCAAAATGACAGTAGATGAGAAAGGTTGTTATATCAGAAATGTTTTCT 27448
Qy 586 TCTCTCTTGAGCCCCCTTCCCTCTGTCCCTCCAGATGGGCGTTGATGCCATAT 645
Db 27449 TAAGCTTTTCTTTCTTTACACCTGCCATGCTCCCAAAATTGGGCATTTAATTCATCT 27508
Qy 646 TTAACCTGGCCATCTCAGTTGCTAAGTGTCTAAGTGTCTTTTCTTTAGGACCCCTTC 705
Db 27503 TTAACCTGGTTTCTGTAGTGTCTAAGTGTCTTTTCTTATAGAACCCCTTC 27568
Qy 706 TTAACGACATATCTGTGACCTGTACTATATAGATCTTTCTGATTAATGCAATTCGGAGATT 765
Db 27569 TGACTGAGCAATATGCTT--CTTGTATTAATAAATCTTTCTGATTAATGCAATGAGGTT 27627
Qy 766 TTTTGTGTAGATGATGAGAGTGGTTCCTGTCTTTTTCACCTTCCTTTACTCAGCTGACTAGT 825
Db 27628 TTTTGTGTAGATGAGAGTGGTTCCTGTCTTTTTCACCTTCCTTTACTCAGCTGACTAGT 27682
Qy 826 GTTCCCTTCTGTTTCTAGTACTGCTGTAGAAATCAGTCTCGGCTTTTACAGTTT 885
Db 27683 GTTCTTCTGTTTCTAGTACTGCTGTAGAAATCAGTCTCGGCTTTTATAGTTT 27742
Qy 886 TAAACTATTTTATGATA-----TTCTGAACATCATCTGTTTTCGACAG 928
Db 27743 TAAATATTTTATGATAATTTCTTAAATATGAACTTTTAAATCATCTGTTTTCGACAG 27802
Qy 929 TACCACACTGTCATGTGATGATGCGGCCCTCTAGACCTCACCACCGGACATG 988
Db 27803 TACCGACACTGTCATGTGACCAATCTG-ACCCTCTTTACCTCGCCGACGACACAG 27861
Qy 989 CTTCCGGTA 997
Db 27862 CCTCCTGTA 27870

RESULT 6

ABK83623

ID ABK83623 standard; cDNA; 1232 BP.

XX

AC ABK83623;

XX

DT 14-AUG-2002 (first entry)

XX

DE Human cDNA differentially expressed in granulocytic cells #194.

XX

KW Human; ss; granulocytic cell; DNA chip; bacterial infection;

KW viral infection; parasitic infection; protozoal infection;

KW fungal infection; sterile inflammatory disease; psoriasis;

KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;

KW cardiac reperfusion injury; renal reperfusion injury; ARDS;

KW adult respiratory distress syndrome; inflammatory bowel disease;

KW Crohn's disease; ulcerative colitis; periodontal disease;

KW granulocyte activation; chronic inflammation; allergy.

XX

OS Homo sapiens.

XX

PN WO200228999-A2.

XX

PD 11-APR-2002.

XX

PF 03-OCT-2001; 2001WO-0821.
XX
XX 03-OCT-2000; 2000US-237189P.
XX (GENE-) GENE LOGIC INC.
PA Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
PI WPI, 2002-435328/46.
XX
XX Detecting granulocyte activation by detecting differential expression
DR of genes associated with granulocyte activation, which serves as
XX diagnostic markers that is useful for monitoring disease states and
XX drug toxicity
XX
XX Claim 1; SEQ ID No 194; 114pp; English.
PS
XX The invention relates to detecting (M1) granulocyte (GC) activation
XX (GCA), by detecting the level of expression of gene(s) (Gs) identified by
XX DNA chip analysis as given in the specification, and comparing
XX the expression level to an expression level in an unactivated
XX GC, where differential expression of Gs is indicative of GCA.
XX Also included are modulating (M2) GA by contacting GC with an agent
XX that alters the expression of at least one gene in Gs. (2) screening (M3)
XX for an agent capable of modulating GCA or an inflammation (especially
XX chronic) in a tissue, an allergic response in a subject, exposure of a
XX subject to a pathogen or sterile inflammatory disease using the
XX gene expression profile; (3) detecting (M4) an inflammation (especially
XX chronic) in a tissue, an allergic response in a subject, exposure of a
XX subject to a pathogen or sterile inflammatory disease, by detecting the
XX level of expression in a sample of the tissue of gene(s) from Gs, where
XX the level of expression of the gene is indicative of inflammation;
XX (4) treating (M5) an inflammation (especially chronic) or in a tissue,
XX an allergic response in a subject, exposure of a subject to a pathogen
XX or sterile inflammatory disease, by contacting a tissue having
XX inflammation with an agent that modulates the expression of gene(s)
XX from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
XX modulating GA; M3 is useful for screening an agent capable of modulating
XX GCA preferably in an inflammation in a tissue; M4 is useful for
XX detecting an inflammation (especially chronic) in a tissue, an allergic
XX response in a subject, exposure of a subject to a pathogen or sterile
XX inflammatory disease (e.g. psoriasis, rheumatoid arthritis, renal
XX glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
XX reperfusion injury, ARDS, adult respiratory distress syndrome,
XX inflammatory bowel disease, Crohn's disease, ulcerative colitis,
XX periodontal disease; also bacterial infection, viral infection,
XX parasitic infection, protozoal infection, fungal infection and M5 is
XX useful for treating one of the above conditions. The present
XX sequence represents a gene differentially expressed in granulocytes.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 1232 BP; 329 A; 240 C; 251 G; 411 T; 1 other;

Query Match 28.0%; Score 285.2; DB 24; Length 1232;
Best Local Similarity 64.4%; Pred. No. 1.7e-78;
Matches 658; Conservative 0; Mismatches 303; Indels 61; Gaps 13;

Qy 8 TCTTGTCTCAGATTGAAGATTGCCAAATACCTTGTGAACATAAGTTGC---GTTGTGCTGA 64
Db 1 TCTTGTCTCAGATTGAAGATTGCCAAATACCTTGTGAACATAAGTTGTGTCGGA 60
Qy 65 GAACAGTAAGCACTAAGCTTTGAGAGACTTTGCTTAAAGAGACTGAGCTTCTGGG 124
Db 61 GAACACCGAGCACTGAACCTTAGCAAGACCTTCTTTCGAGAGAGAGGCTTCTGCA 120
Qy 125 CTCAGG-GGTGCAGACCCCTCCCGTAGC-----TCCAGACCCGTGTGACACAGCAGC 176
Db 121 GTTAGGAGGTGCAGACACTTGTCTCTCTATGTAGTTCTCAGATGCGTAAAGCAGACAGC 180
Qy 177 CTCCTTAAATGACAGCGTGCATGTAAACGACCTGTAACTTATCAGCCCATGCTCATACG 236


```
Db 181 CTCCGAATGAAGCGTTCGCAATGAACACACAGTGAAGTACAGACAGCTGTTCCGACA 240
QY 237 TAACCTTGTACTGTACGTACAGATGGGTGTACAGCTGTCTCTTTGATTTCAATAGTAG 296
Db 241 TAACATTTGATGTATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 300
QY 297 TTCTCTAAATACCAGCTGACCGGCTTCTGAGGCTTTGAACAGAGACTGCTGCTCTGTG 356
Db 301 TTTTCATAGCGAGTTTTCGACCGCTTTTGGAGATTTTGAACAGAACTG-----CTA 354
QY 357 TTGCTCTAACGAAGTATTTCTGTTCTCTAGTGTGGTGTCTGCTGGGTGAGTGTGGAAC 416
Db 355 TTTCTCTTAATGAAGATTTCTGTT--TAGCTGTGGTGTGCGGGTGGGGTGT----- 407
QY 417 ACGACGTCATCAAAAGGACAGACAGATTTTGAC--TAATATGAAGTGAAGATTAATTA 475
Db 408 -----GTGATCAAGAGCAAGACAGATTTTGAACAAATACGAAGTGGAG----ATT 458
QY 476 CACTACATTTGACATGGAGT---AATTCACTGAATAAAGTGTCAAGGTAAAGCTTTT 532
Db 459 CACTACATTTGACAGGAATGAAGTGTCAAGGTAAAGTGTCAAGGTAAAGTGTAAATTC 518
QY 533 TAAAGGTAAATTTCTGTCACACAGATGATGACAAATGCGGATCTTATCAGTGTCTCT 592
Db 519 TCAATATGAGTATGATGAAGAAAGGTGCTTATATCAGGAATGTTTTTAAAGCTT 578
QY 593 TGAGCCCTTCTCCCTGCTGCTCCCTCCAGATGGGGTTCAGTCCATATTAAGTAACT 652
Db 579 TTCTCTTCTTACACCTGCTGCTCCCTCCAGATGGGGTTCAGTCCATATTAAGTAACT 638
QY 653 GGCCATCTCAGCTTCTTAACTTAGCAAGTGTCTTTCTTTAGGACCCCTTCTTAACTGA 712
Db 639 GGTGTTCTGTTAGTCTGCTTAACTTAGTGAAGTGTCTTTCTTATAGAACCCCTTCTG 698
QY 713 GCAATATGCTGACCTGCTATTAAGATCTTTCTGATTAATGCAATTCGGAGATTTTGG 772
Db 699 GCAATATGCTGCTGCTATTAAGATCTTTCTGATTAATGCAATTCGGAGTTTTGGT 757
QY 773 TAGATAGTGAAGTGGTCTGCTTTCTGCTTCTTCTGCTTCTTCTGCTGCTGCTGCTG 832
Db 758 CGATTAGTGAAGTGGTCTGCTTTCTGCTTCTTCTGCTTCTTCTGCTGCTGCTGCTG 812
QY 833 TTGCTTTCTGATTAAGTGGTGTAGAAATCAGCTGCTGCTGCTTCTGCTTCTGCTTCT 892
Db 813 TTAGTTTCTGATTAAGTGGTGTAGAAATCAGCTGCTGCTGCTTCTGCTTCTGCTTCT 872
QY 893 TTTTAGATA-----TTCTGAACATCACTGCTTCTGCTGCTGCTGCTGCTGCTGCTG 935
Db 873 TTTTAGATAATCTTAACTATGAAGCTTCTTAACTCACTGCTTCTGCTGCTGCTGCTG 932
QY 936 ACTGCTATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 995
Db 933 ACTGCTATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 991
QY 996 TA 997
Db 992 TA 993
```

RESULT 7

```
ABN95134
ID ABN95134 standard; DNA; 1232 BP.
AC
XX
AC ABN95134;
XX
DT 13-AUG-2002 (first entry)
XX
DE Gene #1632 used to diagnose liver cancer.
XX
KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW metastatic liver tumour; cytostatic; expression profile; disease state;
KW disease progression; drug toxicity; drug efficacy; drug metabolism.
```

```
XX OS Homo sapiens.
XX PN WO200229103-A2.
XX PD 11-APR-2002.
XX PF 02-OCT-2001; 2001WO-US30589.
XX PR 02-OCT-2000; 2000US-237054P.
XX PA (GENE-) GENE LOGIC INC.
XX PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX WPI; 2002-426119/45.
XX PT Diagnosing and detecting the progression of liver cancer,
XX PT hepatocellular carcinoma or metastatic liver tumor in a patient,
XX PT involves detecting the level of expression of two or more genes in a
XX PT liver tissue sample -
XX PS Claim 1; SEQ ID NO 1632; 298pp; English.
XX CC The invention relates to a novel method for diagnosing and detecting the
XX CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
XX CC tumour in a patient, and differentiating metastatic liver cancer from
XX CC hepatocellular carcinoma in a patient, involving detecting the level of
XX CC expression of two or more genes represented in ABN95134-ABN97455 in a
XX CC tissue sample. The method of the invention has hepatotropic, and
XX CC cytostatic activity. The method is useful for diagnosing and detecting
XX CC the progression of liver cancer, hepatocellular carcinoma and metastatic
XX CC liver carcinoma in a patient. The method is useful for identifying
XX CC expression profiles which serve as useful diagnostic markers as well as
XX CC markers that can be used to monitor disease states, disease progression,
XX CC drug toxicity, drug efficacy and drug metabolism.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 1232 BP; 329 A; 240 C; 251 G; 411 T; 1 other;
```

```
Query Match 28.0%; Score 285.2; DB 24; Length 1232;
Best Local Similarity 64.4%; Pred. No. 1.7e-78;
Matches 658; Conservative 0; Mismatches 303; Indels 61; Gaps 13;
QY 8 TCTTGTCTCAGATTAAAGATTGCCAAATACCTTGTGAAGTGTGCTGCTGCTGCTGCTGCTG 64
Db 1 TCTTGTCTCAGATTAAAGATTGCCAAATACCTTGTGAAGTGTGCTGCTGCTGCTGCTGCTG 60
QY 65 GAACACGTAAAGCACTAAGCTGTTGAGAGACTTTGTCTTAAAGAGACTGCGAGCTTCTGGG 124
Db 61 GAACACCGAGCACTGAACCTTAGCAAGACTTCTGCTTTGAGAGAGCTGAGCTTCTGCA 120
QY 125 CTCAGG-GGTGAGACCCCTCCCGTAGC-----TCCAGACCGGTGTGACAGACAGCAGC 176
Db 121 GTTAGAGGTGCAGACACTTGTCTCTCTATGTAGTTCTCAGATGCTTAAAGCAGACAGC 180
QY 177 CTCCTTAATGACCGCTGCCATGTAAAGCACTGTAACTTATCAGCCCATGCTCATACG 236
Db 181 CTCCTGAATGAAGCGTTGCCATTGAACCTCACCAGTGAAGTGAAGCAGCAGCTGTTCCGACA 240
QY 237 TAACTTTGTACTGTACTGCTCAGCTGCGGTGTAAACGCTCTGCTCTTTGATTTTCATAGTAG 296
Db 241 TAACTTTGTACTGTACTGCTGCGGTGTAAAGCAGCTGAGCTCTTTGGATCAGCTTTGTGA 300
QY 297 TTCTCTAAATACCAGCTGACCGGCTTCTGAGGCTTTGAACAGAACTCTGCTCTGTG 356
Db 301 TTTTCATAGCGAGTTTTCGACCGCTTTTGGAGATTTTGAACAGAACTG-----CTA 354
QY 357 TTGCTCTTAAGCAAGTATTTCTGTTCTTCTAGTGTGGTGTGCTGCTGCTGCTGCTGCTG 416
Db 355 TTTCTCTTAATGAAGATTTCTGTT--TAGCTGTGGTGTGCGGGTGGGGTGT----- 407
```

QY 417 ACAGCGTCATCAAGGACAGACAGCAGTATTTTGAC-TAATATGAAGTAGAGATTAATTA 475
 Db 408 -----GTGATCAAGGACAAAGACAGTATTTTGACAAATACGAAGTGGAG-----ATTTA 458
 QY 476 CACTACATGTACATGGAGT---AATCAACTGAATAAAGTGTACAGGGTAAAGCTTTT 532
 Db 459 CACTACATGTACAGGAATGAAGTGTACAGGGTAAAGTGTACAGGGTAAAGTGTATCTG 518
 QY 533 TAACGGTTAAATTTCTGTCAACAGTAGATGACAAATGCGCGATCTTATCAGTGTCTCT 592
 Db 519 TCAATGAGTAGATGATGAAGAAGTGTGATTAATCAGGAATGTTTCTTAAGCTT 578
 QY 593 TGAGCCCCCTTCCCTGCTGCTCCCTCCAGATGGGGGTGAGTCCATATTTAACT 652
 Db 579 TTCCTTTCTTTACACCTGCCATGCTCCCAATTTGGCAATTTAATCACTTTTAACT 638
 QY 653 GGCCATCCTCAGAGTGTCTAACTAGCAAGTCTTTCTTTAGGACCCCTTCTTAACGA 712
 Db 639 GGTCTTCTGTAGTCGCTAACTTAGTAGTGTCTTTCTTTATAGAACCCCTTCTGACTGA 698
 QY 713 GCAATATGCTGACCTGTACTATAAGATCTTTCTGATAATGCAATTCGGAGATTTTGTGG 772
 Db 699 GCAATATGCTCCT-TCTGTATTAATAAATCTTCTGATAATGCAATTCGGAGATTTTGT 757
 QY 773 TAGATAGTAGAGTGGTTCCTGTTTTCACCTTCTTACTCAGCTGACTAGTCTTCCC 832
 Db 758 CGATTAGTAAAGTGTCTTCCATGTATC-----TTTATTCAGAGCTAATAAGTGTCTTCC 812
 QY 833 TTCGTTTCTAGTAACTGGTGTAGAAATCAGTGTGCGGCTTTTACAGTCTTTTAACTA 892
 Db 813 TTAGTTTCTTAGTAACTAGGTGTAATAATCATGTGTGCGGCTTTTATAGTTTAAATA 872
 QY 893 TTTTATAGTA-----TTCTGAAACATCACTGTCTTGCAGAGTACCAAC 935
 Db 873 TTTTATAGTAATCTTAACTATGAACTTCTTAACATCACTGTCTTGCAGATTAACGAC 932
 QY 936 ACTGTCATGTATGATGATGCGCCCTCTAGACCTCACCACGCGGACATGCTTCCGG 995
 Db 933 ACTGTCATGTATGATGATGCGCCCTCTAGACCTCACCACGCGGACATGCTTCCGG 991
 QY 996 TA 997
 Db 992 TA 993

RESULT 8
 ABL38065/c
 ID ABL38065 standard; cDNA; 605 BP.
 XX
 AC
 XX
 XX
 DT 08-APR-2002 (first entry)
 DE Human colon tumour antigen polynucleotide SEQ ID NO:1654.
 XX
 KW Human; colon cancer; colon tumour antigen; cytostatic; vaccine;
 KW colon tumour metastatic antigen; diagnosis; gene; ss.
 XX
 OS Homo sapiens.
 XX
 FN WO200196388-A2.
 XX
 PD 20-DEC-2001.
 XX
 PF 08-JUN-2001; 2001WO-US18557.
 XX
 PR 09-JUN-2000; 2000US-2108992.
 XX
 PR 20-FEB-2001; 2001US-270216P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Jiang Y, Harlocker SL, Secrlist H;

XX WPI; 2002-114514/15.
 DR Novel isolated colon tumor polynucleotide differentially expressed in
 XX colon tumor or colon metastatic tumor and polypeptides encoded by them,
 PT useful for inhibiting development of cancer in patient -
 PT
 XX Claim 1; SEQ ID 1654; 105pp; English.
 XX ABL36412 to ABL38645 represent human colon tumour antigen cDNA clones (I)
 CC which were isolated from human colon tumour and colon metastatic tumour
 CC cDNA libraries. (II) have cytostatic activity and can be used in vaccine
 CC production. (I) can be used for stimulating and/or expanding T cells
 CC specific for a tumour protein on contact with the T cells. They are also
 CC useful for inhibiting the development of cancer in a patient. (I) can be
 CC used as probes or primers for nucleic acid hybridisation, for preparing
 CC mutant species primers, or primers for use in genetic constructions. (I)
 CC can be used in the diagnosis of a colon tumour.
 XX
 SQ Sequence 605 BP; 219 A; 109 C; 117 G; 156 T; 4 other;
 Query Match 19.0%; Score 193.2; DB 24; Length 605;
 Best Local Similarity 70.5%; Pred. No. 7.7e-50;
 Matches 371; Conservative 0; Mismatches 115; Indels 40; Gaps 7;
 QY 505 GAATAAAGTGTACCGGTAAAGCTTTTAAAGGTTAATTTCTGTCAA--CAGTAGATG 562
 Db 592 GAATGAAGTGTACCGGTAAAGCTTTTAAAGGTTAATTTCTGTCAAATGCGTAGATG 533
 QY 563 ACAAT-----GGCCGATCTTATCAGTGTCTCT-----CTTGAGCCCCCTTCCC 608
 Db 532 ATGAAGAAAGGTGTGTATATATCAGAAATGTTTCTTAAGCTTTCTTCTTCTTACAC 473
 QY 609 CTGCTGTCTCCCTCCAGATGGGGGTGAGTCCATATTTAACTGCGCATCTCCACAGTT 668
 Db 472 CTGCCATGCTCCCAAAATTTGGGCTTTAATTCATCTTTAACTGTTGTTCTGTAGTC 413
 QY 669 GCTAATCTAGCAAGTGTCTTTCTTTAGGACCCCTTCTTTAAGCAATATGTTCTGACCT 728
 Db 412 GCTAATCTAGTAAAGTGTCTTTCTTTAAGCAATATGTTCTGACGAGCAATATGCTT 354
 QY 729 GTACTATAAGATCTTTCTGATTAATGCAATCGGAGATTTTGGTAGATAGTAGAGTGC 788
 Db 353 GTATTATAAATCTTTCTGATAATGCAATAGAGGTTTCTTTGTCGATTAGTAAAGTGC 294
 QY 789 GTTCTGTGTTTTCACCTTCTTCTTACTCAGTGTCTGCTTCTGCTTCTGCTTCTAGTAAC 848
 Db 293 TTTCCAGTTAC-----TTTATTCAGAGCTAATAGTGTCTTCTTCTTCTAGTAAC 239
 QY 849 TGGGTGTAGAAATCAGTGTCTGCGGCTTTTACAGTTTAACTATTTTATGATA----- 901
 Db 238 TAGGTGTAAATCATGTGTGTCAGCTTTATAGTTTAAATATTTTATGATAATTTCTTA 179
 QY 902 -----TTCTGAAACATCACTGTCTTGCAGAGTACCAACATGTCATCTGATTGA 951
 Db 178 AACTATGAACCTTCTTTAATCATCACTGTCTTGCAGATTTACCGACATGGCATTGACCAA 119
 QY 952 TGCCTGCTTCTGACCTTCTAGACCTCACCACGCGGACATGCTTCCGGTA 997
 Db 118 TACTG-ACCCTTTTACCTCGCCACGCGGACACACGCTTCTTCTGTA 74
 RESULT 9
 ABL16478/c
 ID ABL16478 standard; DNA; 605 BP.
 XX
 AC
 XX
 XX
 DT 14-JUN-2000 (first entry)
 DE Human colon cancer differentially expressed nucleotide sequence #483.
 XX
 KW Colon cancer; detect; differential expression; human; treatment;

detect mutation; non-invasive diagnostic method; ds.

Homo sapiens.

WO200012702-A2.

09-MAR-2000.

30-AUG-1999; 99WO-US19424.

31-AUG-1998; 98US-0098639.

27-JAN-1999; 99US-0117393.

(FARB) BAYER CORP.

Endege WO, Steinmann KE, Astle JH, Burgess CC, Carroll E;
Catino TJ, Dwivedi P, Ford DM, Lewis ME, Molino GA, Monahan JE;
Schlegel R;

WPI; 2000-256641/22.

Novel nucleic acids and proteins for identifying therapeutic agents useful for treating and diagnosing cancer, especially colon cancer

Claim 16; Page 319; 3450p; English.

This sequence represents a human nucleotide sequence which is differentially expressed in colon cancer cells compared to the expression levels in normal cells. The nucleotide sequence can be used as a source of primers and probes. The nucleotide sequence is useful for determining the phenotype of a cell by detecting the differential expression of the sequence relative to a normal cell. The probes derived from the sequence can also be used to determine the phenotype of cells in a sample. Probes and antibodies which hybridise to the nucleotide sequence can also be used to determine the phenotype of a cell. The primers are useful for detecting a mutation in a test nucleotide sequence and also for detecting cancer, preferably colon cancer. Antibodies against the protein encoded by the nucleotide sequence can also be used in a method to detect colon cancer. The diagnostic method is non-invasive and accurate for diagnosing colon cancer at an early stage.

Sequence 605 BP; 214 A; 105 C; 116 G; 155 T; 15 other;

Query Match.	18.4%;	Score 187.2;	DB 21;	Length 605;
Best Local Similarity	65.5%;	Pred. No. 5.9e-48;		
Matches 348;	Conservative 0;	Mismatches 158;	Indels 25;	Gaps 5;

QY	485	GTACATGGAGCTAATTCAACTGAATAAAGTGTACGGGTAAAGCTTTTAAACGGTTAATT	544
Db	597	GGACAANGAANNAAGTCGCGGGTAAACCTTTAAANGGTTAAATTTTGTCAAATNCAGTA	538
QY	545	TCTGTCAAACAGTAG-ATGACAAATGGCCGATCTTATCAGTGTCTCTCTTGAGCCCCCT	603
Db	537	GATAAANAANAAGGTTTGNATAAACAGNAATGTTTCTTANGCTTTTCNTTTNTCTT	478
QY	604	TCGCCCTGTCTCCCTCCCGAGATGGGGGTGAGTCCATATTAAACTGSCCATCTCTCA	663
Db	477	AACACCTGCCATCGNTCCCCAAATGGGGCATTTAAATTCATCTTTAAACNGGTTCTTCTGT	418
QY	664	CAGTTGCTAACTTAGCAAGTCCTTTTCTTTAGGACCCCTTCTTAAACGAGCAATATGCT	723
Db	417	TAGTCGCTAACTTAGTAAGTGCTTTTCTTATAGAACCCCTTCTGACTGAGCAATATGCCCT	358
QY	724	GACCTGTACTATAAGATCTTCTGATAATGCATTCGGAGATTTTGTGGTAGATAGTAGA	783
Db	357	-CCTTGTTATTAAATCTTCTGATAATGCATTAGAAGGTTTTTTTGTGATTAGTAAA	299
QY	784	AGTGGGTCCTGTTTTCAACCTTCCTTTACTCAGCTGACTAGTGTCTCCCTTCGTTTTCTA	843
Db	298	AGTGCTTTCCATGTTAC-----TTTATTACAGAGCTAATAGTGCTTTCCTTAGTTTCTA	244
QY	844	GTAACCTGGGTGTAGAAATCAGTGTCTCGGGCTTACAGTTTTTAAACTATTTTAGATA--	901

CC improving cattle. The present sequence is one of the 15112 bovine
CC IMFED EST (expressed sequence tag) nucleic acids.
CC Note: The present sequence was not shown in the specification but
CC was obtained in electronic format from the USPTO web site:
CC seqdata.uspto.gov/sequence.html?DocID=20020137139.
XX
SQ Sequence 432 BP; 123 A; 76 C; 78 G; 155 T; 0 other;

Query Match 16.9%; Score 172.2; DB 25; Length 432;
Best Local Similarity 72.2%; Pred. No. 2.4e-43;
Matches 294; Conservative 0; Mismatches 103; Indels 10; Gaps 5;

QY 486 TACATGGAGTAATCACTGATGAATAAAGTGTACCGGTAAAGCTTTTAAACGGTTAATTT 545
DB 27 TACACGTAAACATTAATACTATTAAAGCATCACGGTAAAGTTTAAAGGTTAATTT 86

QY 546 GTGTCAAAA--CAGTAGATGACAAATGGCGGATCTTTATCAGTGTCTCTTGAGCCCC--- 600
DB 87 GTGTCAAAATGGGTAGATGAAGAAAGGTGGTATTATCAGTGAGTGTCTTTAAAGCTTT 146

QY 601 -CCTTCCCGCTGTCTGCTCCCTCCAGATGGGCGGTGATGTCATATTAACTGGCCATC 659
DB 147 TCCCTTAAACCTCCCAACCCCTTCTGAAATGGGCATTTAATCATCTCTGACCTGTCATC 206

QY 660 CTCACAGTTGCTAACTAGCAAGTGTCTTTCTTTAGGACCCCTCTTAAACGAGCAATAT 719
DB 207 CTCATAGTGTCTAACTAGTAAGTGTCTTTCTTTAGAACTCATCTTTAATGAGCAATAT 266

QY 720 GTCTGACCTGTACTATAGATCTTTCTGATAATGCAATTCGGAGATTTTTTGGTAGATAG 779
DB 267 GCCT-CCITGTATTATAAATCTTCTGATATGCAATAGAAAATTTTTTGTAGTTAG 325

QY 780 TAGAGTGGGTCCTGTTTTCACCTTCCTTACTC--AGCTGACTAGTGCCTCCCTTCT 837
DB 326 TAAAAGTGCACCTCTGTTTTCATGTTCTTTATTCAAAGCTAATAAGTGTCTTCTCTAGT 385

QY 838 TTCTTAGTAACT-GGGTGTAGAAATCACGTGCTGGGGCTTTTACAGTT 883
DB 386 TTCTTAGTAACTAGGTTATAAATAATCACGTGTGGAGCTTTATGGTT 432

RESULT 11
AAS65569
ID AAS65569 standard; cDNA; 1318 BP.
AC AAS65569;
XX
XX
XX
XX 13-FEB-2002 (first entry)
DE
DE DNA encoding novel human diagnostic protein #1373.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
PN
XX
XX 11-OCT-2001.
PD
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
PI
XX WPI; 2001-639362/73.
DR
XX P-PSDB; ABG01382.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 1; SEQ ID No 1373; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1318 BP; 346 A; 268 C; 279 G; 424 T; 1 other;

Query Match 16.5%; Score 168; DB 23; Length 1318;
Best Local Similarity 58.1%; Pred. No. 9.4e-42;
Matches 615; Conservative 0; Mismatches 365; Indels 78; Gaps 15;

QY 1 CCCCTATTTCTGCTCAGATTAGAATGCCAAATACCTTGTGAACATAAGTGC---GTT 57
DB 39 CCCCCATTTCTGTTCAAGATTAAAGATTGCCAAATACCTTGTGAACATAAGTGC 98

QY 58 GTGCTGAGAACACGTAAGCACTAAGCTGTGTGAGAGACTTTGTG-CCTAAGAGACACTGCAG 116
DB 99 GTGCCGAGAACCCGAGCACTGAATTTGCAAGACCTTCGTCTTTGAGAGACCGTAG 158

QY 117 CTTCTGGGCTCAGG-GGTGCAGACCCCTCCCGTAGC-----TCCAGACCGGTGTGACA 167
DB 159 CTTCTGAGTTAGGAGGTGCAGACACTTTTGTCTCTCTATGTAGTTCTCAGATCGGTAAAG 218

QY 168 CAGCAGAGCTCTTAAATGACAGCGTGCATGTAAAGCACTGTAACTTA-TCAGCCCAT 226
DB 219 CAGAACAGCTCCCGAATGAAGCGTTGCCAATTAAGTGTGAACCTACCGAGTGTACGAGCACGT 278

QY 227 GCTCATTAGTAACCTTTGACTGTAGCTCAGATGGGTGTAAACAGCTCTGCTCTTTGATT 286
DB 279 GTTCCGCAATACATATTGCTGTATATGA--GTGAGCGTAGCAGCTCAGCTCTTTGGA 336

QY 287 TCATAGTGAAGTTCTTAAATATACCAGCTGACCGGCTTCTGAGGCTTTTGACAGAACTCT 346
DB 337 TCAGTCTTTGTGATTTTCATAGCGAGTTTCTGACCAGCTTTTCGGGAGATTTTGAACAG 396

QY 347 GGCTCCTGTGTTGCCTCTAACGAGTATTCTGTCTCTAGTCGTTGGGTGTCTGGTGGAG 406
DB 397 AACTGGCTATTTCTCTTAATGAAGGAATTTCTGGTTAGCTGTGGGTGTGCCGGTGGGG 456

QY 407 TGTGTGAACACGACGCTCATCAAGGAGACAGACAGTATTTTCTGACTAATATGAAGTAGAG 466
DB 457 GGTGTGGTGGTGG-----ATCAAGGACAAAGACAGTAAATTT---TGGACAAAATACCG 508

QY 467 ATTAATTTTACATACATTTGTACATGAGTAATTCAACTGAAATAAAGTGTACCGGGTAAA 526
DB 509 AGTGGGAGGATTTTACCACCTACCATTTGGTACAGAGGAATGAAAGTGTGTACCGGGTAAAA 569

QY 527 GCCTTTTAAAGGTTAATTTCTGTCAACAGTATGACAAATGCCCATCTTTATCAGTGT 586
DB 569 CTCCTAAAAGGGTTAATTTCTGTGTCATGCAAGTAGATGAATGAAGAAAGGTTTGTAT 628

```
QY 587 -----CTCTCTGAGCCCTTCCCTGCTGCTCCCTCCCA 624
Db 629 TATCAAGGAATGTTTCTTAAGCCTTTCTCTTCTTCAACCTGCGATGCCCTCCCC 688
QY 625 GA---TGCGGCGTTGAGTCCATATTAACTGGCCATCCTCACAGTTCCTAACTTAGCAA 681
Db 689 AAATTGGGGCATTTAATTATCATGCTTTAACTGGTTGTTCTCTTAGTCGCTTAACCTAGTA 748
QY 682 GTG-----CTTTCTTAGGACCCCTTCTTAAGCAGCAATATGTCGACCTGACTATATA 736
Db 749 AAGGTGCTTTCCCTAAAGAACCCCTTCTGACTGAGCAATATGCTT-CCTTGTATTATA 807
QY 737 AGATCTTCTGATAATGATTCGAGATTTTTTTGGTAGATAGTAGAGTCGCTCTCTGT 796
Db 808 AATCTTCTGATATGATATAGAGGTTTTTTTGTGCGATTAGTAAAGTCCTTCCATG 867
QY 797 TTTACCTCTCTTACTCAGTCACTAGTGTCTCCCTTCTGTTTCTAGTAACTGGGTGA 856
Db 868 TTACTT-----TATTCAGAGCTAATAGTGTCTTCCCTAGTTTCTAGTAACTAGGTGA 922
QY 857 GAATCAGTGTGCGGCTTTACAGTTTAACTATTTAGATA----- 901
Db 923 AAAATCATGTGTTGCAGCTTTATAGTTTAAATAATTTTAAATAATTTTAAACTATGA 982
QY 902 --TCTGAAACATCAGTCTTGCAGAGTACCAACACTGTCATGTCATGATGCGGCC 959
Db 983 ACCTTCTTACATCAGTCTTGCAGATACCGACACTGTCACCTGACCAATACTG-AC 1041
QY 960 CTTCTAGACTACCCACGCGGACACATGCTTCCGGTA 997
Db 1042 CTTCTTTACCTGCGCCACGCGGACACACGCGCTCTCTGTA 1079

RESULT 12
ABZ35389
ID ABZ35389 standard; cDNA; 2167 BP.
XX
AC ABZ35389;
XX
XX 05-FEB-2003 (first entry)
XX
XX Human gene expression profile polynucleotide SEQ ID NO 500.
XX
XX Human; artery; endothelium; umbilical; vein; aorta; pulmonary artery;
XX bronchial epithelium; prostate; muscle; lung fibroblast; osteoblast;
XX tumour; microarray; genome mapping; antibiotic; antiviral; antifungal;
XX gene expression; gene; ss.
XX
XX Homo sapiens.
XX
XX WO200274979-A2.
XX
XX 26-SEP-2002.
XX
XX 20-MAR-2002; 2002WO-US08456.
XX
XX 20-MAR-2001; 2001US-276947P.
XX
XX (ORTH ) ORTHO CLINICAL DIAGNOSTICS INC.
XX
XX Wan J, Wang Y;
XX
XX WPI; 2002-740862/80.
XX
XX New gene expression profile generated from primary, endothelial,
XX epithelial, and muscle cell types, useful for identifying disease
XX pathologies involving alterations of gene expression, e.g. cancer
XX
XX Disclosure; Page 641-642; 850pp; English.
XX
XX The invention relates to a gene expression profile comprising one or more
XX genes (ABZ34889-ABZ35692) and generated from a cell type. The cell type
```

```
CC is a coronary artery endothelium, umbilical artery or vein endothelium,
CC aortic endothelium, dermal microvascular endothelium, pulmonary artery
CC endothelium, myometrium microvascular endothelium, keratinocyte
CC epithelium, bronchial epithelium, mammary epithelium, prostate
CC epithelium, renal cortical epithelium, renal proximal tubule epithelium,
CC small airway epithelium, renal epithelium, umbilical artery smooth
CC muscle, neonatal dermal fibroblast, pulmonary artery smooth muscle,
CC aortic fibroblast, neural progenitor cells, skeletal muscle, astrocytes,
CC bronchial smooth muscle, uterine smooth muscle, lung fibroblast.
CC osteoblasts or prostate stromal cell. The gene expression profile is used
CC for determining the level of RNA expression for a sample, determining the
CC phenotype of a cell and distinguishing cell types. The gene or a protein
CC expression profile is useful in identifying disease pathologies
CC involving alterations of gene expression. The assessment of expression
CC profiles may provide meaningful information with respect to tumour type
CC and stage, treatment methods, and prognosis. The gene or protein
CC expression profile may also be used for creating microarrays. The
CC microarray is useful for genetic and physical mapping of genomes, DNA
CC sequencing, genetic or medical diagnosis, genotyping of organisms,
CC confirming cell or tissue identifications and in identifying promising
CC antibiotics, antiviral or antifungal agents.
XX
SQ Sequence 2167 BP; 537 A; 578 C; 454 G; 594 T; 4 other;
```

```
Query Match 16.1%; Score 164; DB 24; Length 2167;
Best Local Similarity 71.2%; Pred. No. 2.2e-40;
Matches 282; Conservative 0; Mismatches 90; Indels 24; Gaps 4;
```

```
QY 575 TCTATCAGTGTCTCTTGTGAGCCCTTCCCTGCTGCTCCCTCCAGATGGGGGT 634
Db 1780 TGTTTCTCTAAGCCTTTTCCCTTCTTACACCTGCCATGCCCTCCCAATTGGCAT 1839
QY 635 TGAGTCCATATTTAACTGGCCATCTCTCACAGTGTCACTTAGCAAGTGTCTTCTT 694
Db 1840 TTAATTCATCTTTAACTGGTGTCTCTGTAGTCGCTAACTTAGTAGTCTTTCTTAT 1899
QY 695 GGACCCCTTCTTAAAGAGCAATATGTCGACCTGACTATAAGATCTTCTGATAATGC 754
Db 1900 AGAACCCCTTCTGACTGAGCAATATGCCT-CCTTGTATTATAAAATCTTCTGATAATGC 1958
QY 755 ATTGCGAGATTTTTTGGTAGATAGTAGAGTGGTTCCTGTTTACCTTCCTTACTC 814
Db 1959 ATTAGAAGGTTTTTTTGTGATTAAGTAAAGTCTTTCCATGTTAC-----TTTATTCAG 2013
QY 815 AGCTGACTAGTGTCTCCCTGCTGTTTCTAGTAACTGGTGTAGAAATCACTGCTGGGC 874
Db 2014 AGCTAATAGTGTCTTCTTCTAGTAACTGGTGTAGTAAATCATGTGTGTCAGC 2073
QY 875 TTACAGTTTTTAACTATTTTAGATATTCTGA-----AACATCACTGT 918
Db 2074 --TATAGTTTTTAAATAATTTTAGATATTCTTAACTATGAACTTCTTAACTACTGT 2131
QY 919 CTTGCCAGAGTACCAACACTGTCATGTGATTATGC 954
Db 2132 CTTGCCAGATTACCGACACTGTCACCTTGACCAATAC 2167
```

```
RESULT 13
ABL71132
ID ABL71132 standard; cDNA; 285 BP.
XX
AC ABL71132;
XX
XX 14-MAY-2002 (first entry)
XX
XX Corn tassell-derived polynucleotide (cdps) SEQ ID NO:506.
XX
XX Corn; corn tassell-derived polynucleotide; cdps; hybrid breeding; CDPs;
XX inheritance; characteristic; growth; development; disease resistance;
XX environmental adaptability; quality; yield; molecular marker;
XX multigene trait; plant breeding; corn tassell; gene; ss.
XX
```

OS Zea mays.
XX US2001051335-A1.
XX 13-DEC-2001.
XX 16-APR-1999; 99US-0294093.
XX 21-APR-1998; 98US-082567P.
XX (LALG/) LALGUDI R V.
XX (ITOL/) ITO L Y.
XX (SHER/) SHERMAN B K.
XX Lalgudi RV, Ito LY, Sherman BK;
XX WPI; 2002-163647/21.
XX Novel purified corn tassel-derived polynucleotide useful for
PT determining altered gene expression, to recover regulatory elements and
PT to follow inheritance of desirable characteristics through hybrid
PT breeding programs -
XX Claim 1; SEQ ID 506; 201pp; English.
XX The present sequence describes a purified corn tassel-derived
CC polynucleotide sequence (cdps) comprising a nucleic acid sequence
CC selected from those given in ABL70627 to ABL76833. The cdps sequences
CC encode corn tassel-derived polypeptides (CDPs). The cdps sequences (I)
CC can be used for determining altered gene expression, to recover
CC regulatory elements and to follow inheritance of desirable
CC characteristics through hybrid breeding programs. (I) are also useful
CC in the evaluation, and alteration of desired characteristics associated
CC with growth and development, disease resistance, environmental
CC adaptability, quality and yield, and as molecular markers for studying
CC inheritance of multigene traits in a plant breeding program. (I) can be
CC used to produce a tassel-specific profile of gene transcription, a
CC transcript image, to clone regulatory elements for use in transformation
CC vectors, to express a polypeptide, to identify, isolate or extend
CC identical or related corn tassel nucleic acid sequences from DNA
CC libraries, in nucleic acid hybridisation or amplification technologies,
CC as query sequences to determine homology of known sequences, as probe
CC for use in Southern or Northern hybridisation, and to identify the
CC presence of and/or to determine the degree of similarity between two
CC (or more) nucleic acid sequences.
XX Sequence 285 BP; 79 A; 46 C; 44 G; 115 T; 1 other;
SQ
Query Match 12.4%; Score 126.4; DB 24; Length 285;
Best Local Similarity 72.7%; Pred. No. 4.4e-29;
Matches 192; Conservative 0; Mismatches 66; Indels 6; Gaps 2;
QY 658 TCTCAGCTGCTACTAGCAAGTGCCTTTCTTTAGGACCCCTCTTAACGACCAAT 717
Db 5 TTCTGTAGTCGCTACTAGTAGAGTCTTTCTTATAGAACCCCTTCGACTGACCAAT 64
QY 718 ATGCTGACCTGCTACTATAGATCTTTCTGATATGATCGAGATTTTGTGTAGAT 777
Db 65 ATGCTCT-CCTGTATTTATAAAATCTTTCTGATAATGATGATTTTGTGCTAT 123
QY 778 AGTAGAAGTGCCTGCTGCTGTTTTCACCTTCCTTACTGCTAGTCTTCCCTTCGT 837
Db 124 AGTAAAGTGCCTTTCATGTTAC-----TTTATTCAGAGCTATAAGTCTTCCCTTAGT 178
QY 838 TTCTTAGTAAGTGGGTGTAGAAATCAGTCTCGCGCTTTACAGTTTAACTATTTTA 897
Db 179 TTCTTAGTAAGTGGGTGTAGAAATCAGTCTCGCGCTTTATAGTTTAAATATTATTAG 238
QY 898 GATATCTGAAACATCACTGCTTT 921
Db 239 ATAATCTAAACATGATGAACCTCT 262

RESULT 14

ABX43616/c
XX ID ABX43616 standard; cDNA; 416 BP.
XX AC ABX43616;
XX DT 21-FEB-2003 (first entry)
XX DE Bovine EST associated with lactation/muscle/fat deposition #8781.
XX KW Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
KW muscle deposition; fat deposition; genome mapping; gene identification;
KW gene analysis; cattle breeding.
XX OS Bos Taurus.
XX US2002137139-A1.
XX 26-SEP-2002.
XX 24-SEP-2001; 2001US-0960352.
XX 12-JAN-1999; 99US-115707P.
XX 11-JAN-2000; 2000US-0480902.
XX (BYAT/) BYATT J C.
XX (MATH/) MATHIALAGAN N.
XX (TAON/) TAO N.
XX (WARR/) WARREN W C.
XX Byatt JC, Mathialagan N, Tao N, Warren WC;
XX WPI; 2003-110599/10.
XX New nucleic acid associated with lactation, and muscle and fat
PT deposition, useful for genome mapping, gene identification and
PT analysis, cattle breeding, or for genetically improving cattle -
XX Claim 2; SEQ ID No 8781; 245pp; English.
XX The invention relates to a purified nucleic acid molecule associated with
CC lactation or muscle and fat deposition (designated LMFD), derived
CC from cattle, and the LMFD nucleic acid can specifically hybridise to a
CC second nucleic acid molecule comprising any of 15112 nucleotide
CC sequences, appearing as ABX34836-ABX49947, or complements of them.
CC Also included are; (1) a transformed cell having a nucleic acid
CC comprising an LMFD nucleic acid linked to a promoter and a 3' non-
CC translated sequence that functions in the cell to cause termination of
CC transcription and addition of polyadenylated ribonucleotides to a 3' end
CC of the mRNA molecule; and (2) determining a level or pattern of a
CC molecule in a bovine cell or tissue comprising: (a) incubating a marker
CC nucleic acid (comprising any of the 15112 nucleic acid sequences or its
CC complement or fragment) with a complementary nucleic acid molecule
CC obtained from the bovine cell or tissue, where hybridisation between the
CC marker nucleic acid and the complementary nucleic acid permits the
CC detection of the molecule; and (b) detecting the level or pattern of the
CC complementary nucleic acid, where the detection of the complementary
CC nucleic acid is predictive of the level or pattern of the molecule.
CC The LMFD nucleic acid is used for determining a level or pattern
CC of a molecule in a bovine cell or tissue. It is useful for genome
CC mapping, gene identification and analysis, cattle breeding, preparation
CC of constructs for use in cattle gene expression, or for genetically
CC improving cattle. The present sequence is one of the 15112 bovine
CC LMFD EST (expressed sequence tag) nucleic acids.
CC Note: The present sequence was not shown in the specification but
CC was obtained in electronic format from the USPTO web site:
CC seqdata.uspto.gov/sequence.html?DocID=20020137139.
XX SQ Sequence 416 BP; 140 A; 77 C; 101 G; 98 T; 0 other;

Query Match 12.1%; Score 123.2; DB 25; Length 416;
Best Local Similarity 78.2%; Pred. No. 5.4e-28;
Matches 186; Conservative 0; Mismatches 48; Indels 4; Gaps 3;

QY 676 TAGCAAGTGTCTTTCTTTAGGACCCCTTCTTAACGAGCAATATGCTCACCTGTACTAT 735
 |||||
 Db 416 TAGTAAGTGTCTTTCTTATAGAACTCATCTTAATAGCAATATCCCTC-CTTGATTATAT 358
 |||||
 QY 736 AAGATCTTTCTGATAATGCATTCGGAGATTTTTTTGGTAGATAGTAGAAGTGGCTTCCTG 795
 |||||
 Db 357 AAAATCTTTCTGAATATGCATTAGAAATTTTTTTGTAGGTAGTAAAGTGCATCTCTG 298
 |||||
 QY 796 TTTTCACCTTCTTTACTC--AGCTGACTAGTGTCTTCCCTTGGTTTCTAGTAAT-GGG 852
 |||||
 Db 297 TTTTCACTGTCTTTATTCAAAGCTAATAAGTGTCTTCTCTAGTTTCTAGTAAGT 238
 |||||
 QY 853 TGTAGAATAACAGTGTCTGGCTTACAGTTTTTAACTATTTAGATATTTCTGAAC 910
 |||||
 Db 237 TATAAAATACAGGTGTTCAGCTTTATGTTTTTAAAGATCTTCAGATATTCCTCAAC 180
 |||||

RESULT 15

AAH22399
 ID AAH22399 standard; DNA; 348 BP.

XX
 AC AAH22399;

DT 22-AUG-2001 (first entry)

XX Human rac1 gene related nucleotide sequence #3.

XX Identification; toxic; hepatotoxic; differential gene expression;

KW NSAID; non-steroidal antiinflammatory drug; ds.

XX Homo sapiens.

XX WO200138579-A2.

XX 31-MAY-2001.

XX 21-NOV-2000; 2000WO-US32049.

XX 22-NOV-1999; 99US-0166923.

PR 18-FEB-2000; 2000US-0183531.

PR 20-NOV-2000; 2000US-0717321.

XX (CURA-) CURAGEN CORP.

XX Gould-Rothberg BE, Dipippo VA, Ramseh TM, Gerwein RW;

XX WPI; 2001-355948/37.

XX Screening hepatotoxic agent comprises contacting test cell population
 PT expressing RISKMARKER or INJURYMARKER with agent, comparing expression
 PT with reference population and identifying difference in expression
 PT levels -

XX Disclosure; Page 8-9; 76pp; English.

XX The present invention describes a method of screening a test agent for
 CC hepatotoxicity. The method comprises: (a) providing a test cell
 CC population comprising a cell capable of expressing one or more nucleic
 CC acid sequences selected from the group consisting of RISKMARKER 1-8
 CC and INJURYMARKER 1-10; (b) contacting the test cell population with a
 CC test agent; (c) measuring expression of one or more of the nucleic
 CC acid sequences in the test cell population; (d) comparing the
 CC expression of the nucleic acid sequence in the test cell population to
 CC the expression of the nucleic acid sequence in a reference cell
 CC population comprising at least one cell whose exposure status to a
 CC hepatotoxic agent is known; and (e) identifying a difference in
 CC expression levels of the RISKMARKER or INJURYMARKER sequences, if
 CC present, in the test cell population and reference cell population.
 CC The method is useful for identifying a hepatotoxic agent. The present
 CC sequence is given in the exemplification of the present invention.

XX Sequence 348 BP; 103 A; 76 C; 67 G; 102 T; 0 other;

Query Match 11.1%; Score 113; DB 22; Length 348;
 Best Local Similarity 100.0%; Pred. No. 7.7e-25;
 Matches 113; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 905 TGAACATCACTGTCTTCCAGAGTACCAACACTGTCTATGTGATTGATGCCGCCCTCT 964
 |||||
 Db 1 TGAACATCACTGTCTTCCAGAGTACCAACACTGTCTATGTGATTGATGCCGCCCTCT 60
 |||||
 QY 955 AGACTCACTCACTGTCTTCCAGAGTACCAACACTGTCTATGTGATTGATGCCGCCCTCT 1017
 |||||
 Db 61 AGACTCACTCACTGTCTTCCAGAGTACCAACACTGTCTATGTGATTGATGCCGCCCTCT 113
 |||||

Search completed: November 23, 2003, 11:09:39
 Job time : 294.886 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 23, 2003, 14:02:08 ; Search time 448.662 Seconds
(without alignments)
9221.962 Million cell updates/sec

Title: US-09-717-321A-2

Perfect score: 1266

Sequence: 1 ttttttttttttttttcaaa.....gagcaagaataggagatttt 1266

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2169961 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	458.8	36.2	2941	14	US-10-198-846-9874
C 2	447.6	35.4	2051	9	US-09-925-302-255
C 3	430.6	34.0	1644	14	US-10-153-668-48
C 4	430.6	34.0	1646	14	US-10-153-668-46
C 5	420.2	33.2	1232	10	US-09-880-107-1632
C 6	284.2	22.4	2313	14	US-10-198-846-12919
C 7	195.2	15.4	594	14	US-10-066-543-2864
C 8	193.2	15.3	602	10	US-09-878-178-1654
C 9	193.2	15.3	602	13	US-10-046-935-1654
C 10	193.2	15.3	602	14	US-10-146-502-1654
C 11	187.8	14.8	549	14	US-10-066-543-2878
C 12	187.8	14.8	567	14	US-10-066-543-2864
C 13	187.2	14.8	605	11	US-09-871-161-483
C 14	185.4	14.6	477	14	US-10-066-543-3022
C 15	172.2	13.6	432	10	US-09-960-352-7612
C 16	164	13.0	2167	12	US-10-101-510-500

17	163.6	12.9	353	10	US-09-960-352-11432
C 18	147.4	11.6	409	14	US-10-066-543-13
C 19	126.4	10.0	285	9	US-09-294-093B-506
C 20	123.2	9.7	416	10	US-09-960-352-8781
C 21	116.4	9.2	3740	11	US-09-764-891-9984
C 22	108.8	8.6	293	14	US-10-066-543-3077
C 23	98.2	7.8	581	10	US-09-998-998-972
C 24	95.4	7.5	201	10	US-09-960-352-4357
C 25	91.6	7.2	565	10	US-09-998-598-296
C 26	85.6	6.8	492	10	US-09-878-178-1503
C 27	85.6	6.8	492	13	US-10-046-935-1503
C 28	85.6	6.8	492	14	US-10-146-502-1503
C 29	82.2	6.5	398	11	US-09-918-995-34569
C 30	68	5.4	588	11	US-09-871-161-1334
C 31	62.2	4.9	426	10	US-09-960-352-7342
C 32	60	4.7	404	10	US-09-960-352-14206
C 33	57.6	4.5	1651	14	US-10-198-846-13019
C 34	52.4	4.1	162	10	US-09-920-300A-1549
C 35	52.4	4.1	162	12	US-10-099-926-1549
C 36	52.4	4.1	162	13	US-10-033-528-1549
C 37	52.2	4.1	424	10	US-09-960-352-11218
C 38	49.4	3.9	469	11	US-09-918-995-13017
C 39	48.6	3.8	336	12	US-09-814-353-4742
C 40	48.6	3.8	336	12	US-09-814-353-11039
C 41	48.2	3.8	277	10	US-09-960-352-12673
C 42	48	3.8	640681	10	US-09-790-988-1
C 43	47.8	3.8	2218	12	US-09-933-767-232
C 44	47.8	3.8	2218	14	US-10-023-282-103
C 45	46.8	3.7	302	12	US-09-814-353-5335

ALIGNMENTS

RESULT 1

US-10-198-846-9874/c
; Sequence 9874, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9874
; LENGTH: 2941
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-198-846-9874

Query Match 36.2%; Score 458.8; DB 14; Length 2941;
Best Local Similarity 68.0%; Pred. No. 1.4e-113;
Matches 895; Conservative 0; Mismatches 337; Indels 84; Gaps 15;

QY	1	TTTTTTTTTTTTTTTCAAGTTC	CAAGACATTTT	TTTTTTTTTTTTTTTATGATTC	CAAGG 60
Db	2481	TTTTTTTTTTTTTTTTTTTT	TTTTTTTTTTTTTTTTTTTT	TTTTTTTTTTTTTTTATGATTC	CAAGG 60
QY	61	ATTATTAGTCATACATCAAA	CACTGCTA	CTGCTAGCAAAAGATCAATG	TAA 120
Db	2421	ATTATTAGTCATACATCAAA	CACTAATG	CTAATGCTAGCAAAAGATCAATG	TAA 2362
QY	121	AAACTCCAAATTCGCAACTG	CAATG	AAAAAGTTTCTAGTGGT	CGAAGGC 180


```
Db 1650 TTAAGAATTACTAAATATTTTAAATACTATAAGCTGCAACATGATTTTACACT 1591
Qy 413 AGTTACTAGAAAACGAAGGAGCACTAGTCACTGAGTAAGAGAGGTGAAACAGAA 472
Db 1590 AGTTACTAGAAAACGAAGGAGCACTATTAGCTCTGAATA-----AAGTAACATGAA 1536
Qy 473 CGCACTTCTACTATCTACCAAAAAAATCTCCGAATGCAATTATCAGAAGATCTTATAGTA 532
Db 1535 AGCACTTTTACTAATCGCAAAAAAACCTTCTAATGCAATTATCAGAAGATTTTATAATA 1476
Qy 533 CAGTCTGACATATGCTCGTTAAAGAGGGGTCTTAAGAAAAAGCACTTCTCTAAGTTAG 592
Db 1475 CAAG-GAGGCATATTGCTCAGTCAGAAGGGTCTTATAAGAAAAGCACTTCTAAGTTAG 1417
Qy 593 CAATCTGTGAGATCGGCAGTTTAATATATGACTCAAGCCCACTCTGGGAGGAGACGA 652
Db 1416 CGACTAAGACAAACACAGTTTAAAGATGAATTAATGCCAATTTGGGAGGATGCA 1357
Qy 653 GGGGAAGGGGGCTCAAGAGAGACACTGATAAGATCGGCATTTGTCATCTACTGTTG 712
Db 1356 GGTGTAGAGAAAGAAAGCTTAAGAAAACATTTCTGTATATACCAACCTTTCTTTCA 1297
Qy 713 ACAGAAATTAACCGTTAAAGCTTTTACCCTGTGACACTTTTATTCAGTTGAA---TTACT 769
Db 1296 TCATCTACTGCATTTGACAGAAATTAACCTTTTAGAGTTTATACCGTGACACTTTTCA 1237
Qy 770 CCAATGACATGATGCTTAATTAATTAATCTCTACTTCTATA-TTAGTCAAAATCTCTGTGC 828
Db 1236 CTTGTACAATGAGTGTGTA-----AATCTCCACTTCGTATTTTGTCAAAATCTCTTTG 1181
Qy 829 TCCCTTTGATGACGTGCTGTTTCAACACTCCACCAGCACACCACCACTAGGAACAGAA 888
Db 1180 TCCCTTTGATCA-----CACACACCCACCCCGCACCCACAGCTA--AACAGAA 1133
Qy 889 TACTTCTGTAGAGCAACACAGGAGCGAGTTCTGTTCAAAGCCTGCAAGACCGGTCA 948
Db 1132 TTCTTCATTAGAGGAATA-----GCAGTTCTGTTCAAATCTCCGCAAGAGCTGTCA 1079
Qy 949 GCTGGTATTTTAGAGACTCACTATGAAATCAAGAGCAGAGCTGTTTACACCCATCGTGA 1008
Db 1078 GAAACTCGGTATGAAATCAAAAGACTGATCAAAAGAGCTGAGCTGCTACGCTCACTCC 1019
Qy 1009 CGTACAGTCAAAAGTTACGTAATGAGCATGGCTGATAGTTACAGTGGGTACATGGC 1068
Db 1018 ATTACAGTCAATGTTATGTCGGGAACAGCTGCTGCTACTCACTGCTGAGTTCAATGGC 959
Qy 1069 AGCGTGTCAATTAAGAGGCTGCTGTGTGCACACGCTGSG-----GAGCTACGGGAG 1121
Db 958 AAGCGTTTCATTCGGGAGGCTGTTCTGCTTTACGCATCTGAGAACTACATAGGAGCAAG 899
Qy 1122 GGTCTGCACC-CCTGAGCCAGAGCTGCAGTCTTCTTAAGGCAAGTCTCTCAACAGC 1180
Db 898 TGTCTGCACCTCTAACTCGAGAAGCTACCGTCTCTCAAAGACGAGGTCTTTGCAAG 839
Qy 1181 TTAGTGTCTACGTGTTCTCAGACAAAC---GCAACTTAGTTTCAAGAGTATTTTGCAT 1237
Db 838 TTCAGTGTCTGGTGTCTCGGCACAAATGCAATGAGTTTCAAGAGGTATTTTGGCAAC 779
Qy 1238 TCTTAATCTGAGCAAGATAGGGG 1261
Db 778 TCTTAATCTGAACAGAAATGGGG 755
```

RESULT 3

US-10-153-668-48/c

; Sequence 48, Application US/10153668

; Publication No. US2003092616A1

; GENERAL INFORMATION:

; APPLICANT: HONDA, Goichi

; APPLICANT: MATSUDA, Akio

; APPLICANT: MURAMATSU, Shuji

; APPLICANT: ISHIZAWA, Kenya

```
; TITLE OF INVENTION: STAT6 Activating Gene
; FILE REFERENCE: 1254-0207P
; CURRENT APPLICATION NUMBER: US/10/153,668
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: US 60/293,172
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/316,031
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/328,403
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: JP 2001-157043
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: JP 2001-260681
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: JP 2001-313175
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 488
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 48
; LENGTH: 1644
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (109)..(447)
US-10-153-668-48
```

```
Query Match 34.0%; Score 430.6; DB 14; Length 1644;
Best Local Similarity 67.8%; Pred. No. 4.4e-106;
Matches 859; Conservative 0; Mismatches 324; Indels 84; Gaps 15;

Qy 50 ATGATTCAGAGATTTATTAAAGTCATACATGCAAAACATCTGCTAACTGCTAGCAAAA 109
Db 1644 ATGATTCAGAGATTTATTAAAGTCATACATGCAAAACATATGCTAACTGCTAGCAAAA 1595
Qy 110 GATCAATGTAAACACTCCACAATTCGCACTGCTGCAATGAAAGTTTGTCTAGT 169
Db 1584 GATCAATGTAAACACTCCACAATTCGCACTGCTGCAATTTAAAAAATCTGTGTAGT 1525
Qy 170 GGTGGAAGGCCCAACACTGTGTTCTTTCAGTGAGTTAGTTGTACAGAACGGGTAG 229
Db 1524 GGTGGAAGGTCCTCAGCTGATTTCTGCGCAGTGAGTTAAGTTGTACAGAACATCGTCAG 1465
Qy 230 CACTAGCGC--TTGACAGAACCTTACAGACCCCAAGG-----TA 266
Db 1464 CACTAGCAGAGTTTACAGAACCTTACAGACCCCAAGAACATCAATAGGCAAGCGACTA 1405
Qy 267 CCGGAAGCATGTCTCCGCTGGGTGAGGTCCTAGAGGGGGGGCATCAATCAGATGACGT 326
Db 1404 CAGGAGGCGTGTCTCCGCTGGGCGAGGTAAGA-GGGTCAGTATTGGTCAAGTGACAGT 1346
Qy 327 GTTGTGACTCTGCAAGACAGATGATG-----TTTCAGAATATCTAAAA 369
Db 1345 GTCCGTAACTCGCAAGACAGATGTTTAAAGAGTTTCAAGTTTAAAGATTTATCTAAAA 1286
Qy 370 TAGTTTAAAAACTGTAAAGCCGACGACGCTGATTTCTACCCAGTCTACTAGAAAAACGAA 429
Db 1285 TATTTTAAAAACTATAAAGCTGCAACACATGATTTTACCTAGTTACTAGAAAACTAA 1226
Qy 430 GGAAGCACTAGTCACTGAGTAAAGGAAGGTGAAAAAGGAGCGCACTTCTACTACTA 489
Db 1225 GAAAAGCACTTATTAGCTCTGAATA-----AAGTAACATGGAAGCACTTTTACTAATCG 1171
Qy 490 CCAAAAAAATCTCCGAATGCAATTATCAGAAGATCTTATAGTACAGGTGACAGATATTGG 549
Db 1170 ACAAAAAACCTTCTAATGCAATTATCAGAAAGATTTTATAACAG-GAGGCATATTGGC 1112
Qy 550 TCGTTAAGAGGGGTCTCTAAAGAAAAAGCACTTGTAGTTAGCAACTCTGAGGATGGCC 609
Db 1111 TCAGTCAGAGGGGTTCTATAAGAAAAAGCACTTCTAAGTTAGCGACTACACAGAACACC 1052
Qy 610 AGTTTAAATATGACTCAACGCCCATCTGGGAGGGGACAGCAGGGGGAGGGGGCTCA 669
```

Db	1051	AGTTTAAGATGAATTAATTCGCCAATTTGGGAGGAGCATGCAGGTGTAAAGAAAGGAA	992
Qy	670	AGAGAGACACTGATAAGATCGGCCATTGTGCATCTACTGTTTGAACAGAAATTAACCGTTA	729
Db	991	AAGCTTAAGAAAACATTTCTCTGATATACCAACCTTTCTTTCATCATCTACTGCAATTGA	932
Qy	730	AAAAGCTTTACCGTGACACTTTTATTCAGTTGAA---TTACTCCATGATCAATGTAGTG	786
Db	931	CAGAAATTAACCTTTTAGAGTTTATACCCGTGACACTTTTCATTCCTGTCAATGTAGTG	872
Qy	787	TAAATTAATCTCTACTTTCATA-TTATGTCAAAATACTGCTGCTCTCTTTTGATGACGTGT	845
Db	871	TA-----AATCTCCACTCTGATTTTGTCAAAATACTGCTTTGTCTCTTGAATCA	822
Qy	846	GTTCACACACTCCACCCAGCACACCCAGCTAGGAAACAGAAATCTTCGTTAGAGGAA	905
Db	821	-----CACACACCCACCCGGCACACCCACAGCTA--AACAGAAATCTTCATTAGAGGAA	768
Qy	906	CACAGGACCGAGATTCTGTTCAAGCCCTGAGAAGCCGGTCAGCTGGTATTTTAGAGAA	965
Db	767	TAG-----CAGTTCTGTTCAAAATCTCCGCAANAAGCTGGTCAGAAAACCTCGCTATGAA	714
Qy	966	CTCACTATGAATCAAGAGCAGAGCTGTTACACCCATCGTGACGTACAGTACAAAGTTA	1025
Db	713	TCACAAAGACTGATCCAAAGAGCTGAGCTGTACGCTACTCCATTACGTACATGTTA	654
Qy	1026	CGTAATGAGCATGGGCTGATTAAGTTACAGTGGTGTACATGCGAGCGTGTCTAATAGGAG	1085
Db	653	TGTCGGAAACAGCTGCTCTAATCTCACTGGTGAGTTCAATGCAACGCTTTCATTCCGGAG	594
Qy	1086	GCTGTGCTGTGTCACACGGTCTGG-----GAGTACGGGAGGCTGTCACC-CCTGAG	1137
Db	593	GCTGTCTGCTTTAGCGATCTGAGAACTACATAGGAGCAAGTGTCTGCACTCTCTAAC	534
Qy	1138	CCGAGAGCTGAGCTCTCTTAAGACAAAGTCTCTCAACAGCTTAGTCTTAAGTGTTC	1197
Db	533	TGCAAGAGCTACCGTCTTCTCAAGACGAAGTCTTTGCAAGTTCAGTGTCTGGTGTTC	474
Qy	1198	TCAGCACAAAC---GCAACTTAGTTTCAAGGATTTTGGCAATTTCTTAATCTGAGCAAGA	1254
Db	473	TCGGCACAAATGCACTGATGTAGTTTCAAGAGGTATTTTGGCAACTCTTAAATCTGAACA	414
Qy	1255	ATAGGGG 1261	
Db	413	ATGGGGG 407	
RESULT 4			
US-10-153-668-46/c			
; Sequence 46, Application US/10153668			
; Publication No. US20030092616A1			
; GENERAL INFORMATION:			
; APPLICANT: HONDA, Goichi			
; APPLICANT: MATSUDA, Akio			
; APPLICANT: MURAMATSU, Shuji			
; TITLE OF INVENTION: STAT6 Activating Gene			
; FILE REFERENCE: 1254-0207P			
; CURRENT APPLICATION NUMBER: US/10/153,668			
; CURRENT FILING DATE: 2002-05-24			
; PRIOR APPLICATION NUMBER: US 60/293,172			
; PRIOR FILING DATE: 2001-05-25			
; PRIOR APPLICATION NUMBER: US 60/316,031			
; PRIOR FILING DATE: 2001-08-31			
; PRIOR APPLICATION NUMBER: US 60/328,403			
; PRIOR FILING DATE: 2001-10-12			
; PRIOR APPLICATION NUMBER: JP 2001-157043			
; PRIOR FILING DATE: 2001-05-25			
; PRIOR APPLICATION NUMBER: JP 2001-260681			
; PRIOR FILING DATE: 2001-08-30			
; PRIOR APPLICATION NUMBER: JP 2001-313175			
; PRIOR FILING DATE: 2001-10-10			
; NUMBER OF SEQ ID NOS: 488			
; SOFTWARE: PatentIn Ver. 2.0			
; SEQ ID NO 46			
; LENGTH: 1646			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; FEATURE:			
; NAME/KEY: CDS			
; LOCATION: (109)..(273)			
US-10-153-668-46			
Query Match			
Best Local Similarity 34.0%; Score 430.6; DB 14; Length 1646;			
Matches 859; Conservative 67.8%; Pred. No. 4.4e-106;			
Mismatches 324; Indels 84; Gaps 15;			
Qy	50	ATGATTCAGGATTTATTAAGTCATACATGCAAAACACTACTGCTAACTGCATTAGCAAAA	109
Db	1646	ATGATTCAGGATTTATTAAGTCATACATGCAAAACATAAATGCTAAATTCATTAGCAAAA	1587
Qy	110	GATCAATCTAAACACACTCCCAATTTCTGCAACTCTCAATTTGAAAAAGTTTGTCTAGT	169
Db	1586	GATCAATGTAAACACTCCCAATTTCTGCAACTCTCAATTTTAAAAAATTCGTGTAGT	1527
Qy	170	GCTCGAAAGGCCCAACACTGTGTTCTTGCCAGTGAAGTTAGTTGTACAGAACGGGTAG	229
Db	1526	GGCTGAAGGGTCCCAACGCTGATTTCTCGCCAGTGAAGTTAGTTGTACAGAACATCGTCAG	1467
Qy	230	CACTAGCCG--TTGACAGAACTCCACAGACCCCAAGG-----TA	266
Db	1466	CACTAGCACAGTTTACAGAACTCCACAGACCCCAAGAACTCAATAGGCAAGCGACTA	1407
Qy	267	CCGGAAGCATGTGTCGGGTGGGTGAGGTCTAGAGGGGGCGGCATCAATCACATGACAGT	326
Db	1406	CAGGAGGCGTGTGTCGGGTGGCGAGGTAAAGA--GGGTCAATGTTGGTCAAGTGACAGT	1348
Qy	327	GTGCTACTCTGGCAAGACAGTGAAGT-----TTTCAGAAATATCTAAAA	369
Db	1347	GTGCTAACTCTGGCAAGACAGTGAAGTGAAGAGGTTTCATAGTTTAAAGAAATATCTAAAA	1288
Qy	370	TAGTTTAAACACTGTAAGCCCGCAGCAGTGAATTTCTACACCCAGTTACTAGAAAAAGAA	429
Db	1287	TATTTTAAACACTAATAAGCTGCAACACATGATTTTACACCTAGTTTACTAGAAAACTAA	1228
Qy	430	GGGAAGCACTAGTCAAGTGAAGTAAAGAGAGGTGAAACAGGAAACCGCACTTCTACTACTA	489
Db	1227	GGAAAGCACTTATAGTCTGAATA-----AAGTAACATGGAAGACACTTTACTAATCG	1173
Qy	490	CCAAAAAATCTCCGATGCAATTCAGAAAGATCTTATAGTACAGGTGACAGATATGCG	549
Db	1172	ACAAAAAACCTTCTAATGCAATTCAGAAAGATTTTATAACAAG--GAGGCATATGCG	1114
Qy	550	TCGTTAAGAGGGGGTCTTAAAGAAAGCACTTGTCTAAGTTAGCAACTGTGAGGATGGCC	609
Db	1113	TCAGTCAGAAAGGGTTCCTAAGAAAGCACTTACTAAGTTAGCGACTACAGAAACAACC	1054
Qy	610	AGTTTAAATATGGACTCAACCGCCCATCTGGGAGGAGCAGACAGGGGGAGGGGGCTCA	669
Db	1053	AGTTTAAAGATGAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG	994
Qy	670	AGAGAGACACTGATTAAGATCGGCCCATTTGTCTACTGTTTTCACAGAAATTAACCGTTA	729
Db	993	AAAGTTTAAAGAAACATTTCTGATAATACCAACCTTTCTTCTATCATCTACTGATTTGA	934
Qy	730	AAAAGCTTTTACCGTGCACACTTTTATTCAGTTTGA---TTACTCCATGTACATGATG	786
Db	933	CAGAAATTAACCTTTTAGAGTTTATACCGGTGCACACTTTCATTCCTTGTACATGATG	874
Qy	787	TAAATTTAATCTCTACTTTCATA-TTAGTCAAAAATCTGTCTGTCTCTCTTGTATGACG	845
Db	873	TA-----AATCTCCACTTCGTTATTTGTCAAAAATCTGTCTCTCTCTTGTATGATCA	824
Qy	846	GTTCACACACTCCACCCAGCACACCCAGCTAGGAAACAGAAATCTTCGTTAGAGGCAA	905
Db	823	-----CACACACCCCAACCCGGCACACCCACAGCTA--AACAGAAATTCCTTATTAGAG	770


```
RESULT 6
US-10-198-846-12919/c
; Sequence 12919, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12919
; LENGTH: 2313
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-198-846-12919

Query Match      22.4%; Score 284.2; DB 14; Length 2313;
Best Local Similarity 63.5%; Pred. No. 2.9e-66;
Matches 651; Conservative 0; Mismatches 318; Indels 57; Gaps 12;

QY 265 TACCGAGACATGTGTCGGTGGTGGTGTAGAGGGGGGGGCAATCAATCACATGACA 324
DB 221 TACAGAGCGGTGTGTCGGTGGGCGAGGTAAAG-GGGTGAGTATGGTCAAGTGACA 2163
QY 325 GTGTTGGTACTCGGCAAGACAGTGTATTTTCAGAA-----TATCTAA 367
DB 2162 GTGTCGTAACTGGCAAGACAGTGTATTAAGAGGTTTCATAGTTTAAAGATTATCTAA 2103
QY 368 AATAGTTTAAACTGTAAGCGGAGAGCGTGTATTTACACCCAGTTACTAGAAAAG 427
DB 2102 AATATTTTAAACTATAAGCTGCAACACATGATTTTACACCTAGTTACTAGAAAAG 2043
QY 428 AAGGAGACACTAGTCAGCTGAGTAAAGAGGTGAAGACAGAAACGACCTTCTCATATC 487
DB 2042 AAGGAAGCACTTATAGCTCTGAATA-----AAGTAACTGGAAGACCTTTACTAAT 1988
QY 488 TACCAAAAAATCTCGAATGCAATATCAAGAAAGATCTTATAGTACAGGTCAGACATAT 547
DB 1987 CGACAAAAAAACCTTCTAATGCATTTATCAAGAAAGATTTTATAATAAAG-GAGGCATAT 1929
QY 548 GCTCGTTAAGAGGGGGTCTTAAGAAAGACCTTCTAGTTAGCAACTGTGAGGATGG 607
DB 1928 GCTCAGTCAAGAGGGGTCTTATAAGAAAGACCTTACTAAGTTAGCGACTAACAGAACAA 1869
QY 608 CCAGTTTAAATATGGACTCAACCCCATCTGGGAGGAGGACAGCAGGGGGAAGGGGGCT 667
DB 1868 CCAGTTTAAAGATGAATTAATCCCAATTTGGGAGGACATGGCAGGTGTAAGAGAGG 1809
QY 668 CAAGAGACACTGATAAGATCGGCCATTTGTCTACTCTGTTTGTGACAGAAATTAACCGT 727
DB 1808 AAAAGCTTAAGAAAAACATTTCTCTGATTAATCAACCAACCTTTCTTCATCATCTACTGCAAT 1749
QY 728 TAAAAAGCTTACCGTGACACTTTTATTCAGTTGATTAATCTCATGTACATGACATGATGT 787
DB 1748 GACAGAAATTAACCTTTTAGAGTTTTTACCCGAGACACT--TTCTTCTTTGACAAATGT 1691
QY 788 AATATTAATCTACTTCTATCATA--TTAGTCAAAATACTGTCTGTCTCTTTGATGACGTCGT 846
DB 1690 AGTGTATCTCCACTTGTATTTTGICAAAATACTGTCTTTTGCTTTGATCA----- 1638
QY 847 TTTCACACACTCCACCCAGACACCCACGACTAGGAACAGAAATCTTCGTTAGAGGAAC 906
DB 1637 ---CACACACCCACCCGGCACACCAAGCTA--AAACAGAAATCTTCTATTAGAGGAAT 1583

US-10-066-543-2864/c
; Sequence 2864, Application US/10066543
; Publication No. US2003008781A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indrias, Carol Yoseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Margarita
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.563
; CURRENT APPLICATION NUMBER: US/10/066,543
; CURRENT FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2864
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-066-543-2864

Query Match      15.4%; Score 195.2; DB 14; Length 594;
Best Local Similarity 70.9%; Pred. No. 1.9e-42;
Matches 373; Conservative 0; Mismatches 113; Indels 40; Gaps 7;

QY 265 TACCGAAGCATGTGTCGGTGGTGGTGTAGAGGGGGGGGCAATCAATCAATGACA 324
DB 524 TACAGAGGGGTGTGTCGGTGGGCGAGGTAAAGA-GGTGAGTATTTGCTCAAGTGACA 466
QY 325 GTGTGCTACTCTGGCAAGACAGTGTATTTTCAGAA-----TATCTAA 367
DB 465 GTGTGCTAATCTCGCAAGACAGTGTATTTAAGAGGTTTCATAGTTTAAAGATTTATCTAA 406
QY 368 AATAGTTTAAACTGTAAAGCCGACAGCTGATTTCTACCCAGTTACTAGAAAAG 427
DB 405 AATATTTTAAACTATATAAGCTGCAACACATGATTTTACACCTAGTTACTAGAAAAG 346
```



```

QY 608 CCAGTTTAAATATGACTCAACGCCCATCTGGGAGGACAGCAGG-----GGG 657
Db 427 CCAGTTTAAAGATGAATTAATGCCCCAATTTGGGAGGATCGCAGGTGTAAGAGAAAG 486
QY 658 AAGGGGGGCTCAAGAGACACTGATAAGATCGGC---CAATTGTCACTACTG---TTT 711
Db 487 AAAAGCTTAAAGAAACAATTTCTGATAATACCAACCTTTCTTTCATCATCTACTGCAATT 546
QY 712 GACAGAAATTAACCGTTAAAGAGCTTTTACCCGTGACACTTTTATTC 757
Db 547 GACAGAAATTAACCTTTTAGAGNTTTTACCNTGACACTTTTCATTC 592

RESULT 10
US-10-146-502-1654
; Sequence 1654, Application US/10146502
; Publication No. US20030069180A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; APPLICANT: Wang, Aijun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.527C2
; CURRENT APPLICATION NUMBER: US/10/146,502
; CURRENT FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2241
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1654
; LENGTH: 602
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 569..578
; OTHER INFORMATION: n = A,T,C or G
US-10-146-502-1654

```

```

Query Match 15.3%; Score 193.2; DB 14; Length 602;
Best Local Similarity 70.5%; Pred. No. 6.7e-42;
Matches 371; Conservative 0; Mismatches 115; Indels 40; Gaps 7;

QY 265 TACCGAAGCATGTGTCCGCGTGGTGAGGTCTAGAGGGGGCGGCATCATCATGACA 324
Db 74 TACAGGAGCGGTGTGTCCGCGTGGCGAGGTAAAGA-GGGTCAGTATTGGTCAAGTGCCA 132
QY 325 GTGTTGGTACTCTGGCAAGACAGTGATGTTTCAGAA-----TATCTAA 367
Db 133 GTGTCGTAATCTGGCAAGACAGTGATGTTTAAGAAGTTTCATAGTTTAAGAAATATCTAA 192
QY 368 AATAGTTTAAAACTCTAAAGCGCAGCAGTGATTTTACACCCAGTTACTAGAAAACG 427
Db 193 AATATTTTAAAACTATAAAGCTGCAACACATGATTTTACACCTAGTTACTAGAAAAC 252
QY 428 AAGGGAAGCACTAGTCAGTCGTAAGGAAGGTGAAGAGCAAGCAAGCACTTCTACTATC 487
Db 253 AAGGAAGCACTTATTAGCTCTGAATA-----AAGTAACATGGAAGCACTTTTACTAAT 307
QY 488 TACCAAAAAATCTCCGAATGCATTATCAGAAAGATCTTATAGTACAGGTTCAGACATATT 547
Db 308 CGCAAAAAAACCTTCTAATGCATTATCAGAAAGATTTATATAACAAG-GAGGCATATT 366
QY 548 GCTCGTTAAGAGGGGGTCTTAAAGAAAGCACTTCTAAGTTAGCACTGTGAGGATGG 607
Db 367 GCTCAGTCAGAGGGGTTCTATAAGAAAGCACTTACTAAGTTAGGCACTTAACAGAACAA 426
QY 608 CCAGTTTAAATATGAGTCAACGCCCATCTGGGAGGACAGCAGG-----GGG 657
Db 427 CCAGTTTAAAGATGAATTAATGCCCAATTTGGGAGGATCGCAGGTGTAAGAGAAAG 486

```

```

QY 658 AAGGGGGGCTCAAGAGACACTGATAAGATCGGC---CAATTGTCACTACTG---TTT 711
Db 487 AAAAGCTTAAAGAAACAATTTCTGATAATACCAACCTTTCTTTCATCATCTACTGCAATT 546
QY 712 GACAGAAATTAACCGTTAAAGAGCTTTTACCCGTGACACTTTTATTC 757
Db 547 GACAGAAATTAACCTTTTAGAGNTTTTACCNTGACACTTTTCATTC 592

RESULT 11
US-10-066-543-2878
; Sequence 2878, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indrias, Carol Yoseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Margarita
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.563
; CURRENT APPLICATION NUMBER: US/10/066,543
; CURRENT FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2878
; LENGTH: 549
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 526
; OTHER INFORMATION: n = A,T,C or G
US-10-066-543-2878

```

```

Query Match 14.8%; Score 187.8; DB 14; Length 549;
Best Local Similarity 73.3%; Pred. No. 1.9e-40;
Matches 304; Conservative 0; Mismatches 87; Indels 24; Gaps 4;

QY 265 TACCGAAGCATGTGTCCGCGTGGTGAGGTCTAGAGGGGGCGGCATCATCATGACA 324
Db 74 TACAGGAGCGGTGTGTCCGCGTGGCGAGGTAAAGA-GGGTCAGTATTGGTCAAGTGACA 132
QY 325 GTGTTGGTACTCTGGCAAGACAGTGATGTTTCAGAA-----TATCTAA 367
Db 133 GTGTCGTAATCTGGCAAGACAGTGATGTTTAAGAAGTTTCATAGTTTAAGAAATATCTAA 192
QY 368 AATAGTTTAAAACTCTAAAGCGCAGCAGTGATTTTACACCCAGTTACTAGAAAACG 427
Db 193 AATATTTTAAAACTATAAAGCTGCAACACATGATTTTACACCTAGTTACTAGAAAAC 252
QY 428 AAGGGAAGCACTAGTCAGTCGTAAGGAAGGTGAAGAGCAAGCAAGCACTTCTACTATC 487
Db 253 AAGGAAGCACTTATTAGCTCTGAATA-----AAGTAACATGGAAGCACTTTTACTAAT 307
QY 488 TACCAAAAAATCTCCGAATGCATTATCAGAAAGATCTTATAGTACAGGTTCAGACATATT 547
Db 308 CGCAAAAAAACCTTCTAATGCATTATCAGAAAGATTTATATAACAAG-GAGGCATATT 366
QY 548 GCTCGTTAAGAGGGGGTCTTAAAGAAAGCACTTCTAAGTTAGCACTGTGAGGATGG 607
Db 367 GCTCAGTCAGAGGGGTTCTATAAGAAAGCACTTACTAAGTTAGGCACTTAACAGAACAA 426
QY 608 CCAGTTTAAATATGAGTCAACGCCCATCTGGGAGGACAGCAGGAGGAGG 662
Db 427 CCAGTTTAAAGATGAATTAATGCCCAATTTGGGAGGATCGCAGGTGTAAGAG 481

```


RESULT 12

US-10-066-543-2964
; Sequence 2964, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indrias, Carol Yoseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Margarita
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.563
; CURRENT APPLICATION NUMBER: US/10/066.543
; CURRENT FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2964
; LENGTH: 567
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 527
; OTHER INFORMATION: n = A,T,C or G
US-10-066-543-2964

Query Match 14.8%; Score 187.8; DB 14; Length 567;
Best Local Similarity 73.3%; Pred. No. 1.9e-40;
Matches 304; Conservative 0; Mismatches 87; Indels 24; Gaps 4;

QY 265 TACCGAGCATGTGTCGGGTGGGTGAGGTCTAGAGGGGGCGGCATCAATCAGATGACA 324
Db 74 TACAGGAGCGGTGTGTCGGGTGGGTGAGGTCTAGAGGGGGCGGCATCAATCAGATGACA 132

QY 325 GTGTTGCTACTCTGGCAAGACAGTGATGTTTCAGAA-----TATCTAA 367
Db 133 GTGTCGTAATCTGGCAAGACAGTGATGTTTCAGAAAGGTTTATAGTTTAAAGAAATTAUCTAA 192

QY 368 AATAGTTTAAAACTGTAAAGCCGACAGCTGATTTCTACCCAGTTACTAGAAAACG 427
Db 193 AATATTTTAAAAACTATAAGCTGCAACACATGATTTTACACCTAGTTACTAGAAAAC 252

QY 428 AAGGAGACACTAGTCAGCTGAGTAAAGGAGGTGAAACAGGAAACGCACTTCTACTATC 487
Db 253 AAGGAAAGCACTTATTAGCTCTGAATA-----AAGTAAACATGGAAGCACTTTTACTAAT 307

QY 488 TACCAAAAAATCTCCGAATGCATTATCAGAAAGATCTTATAGTACAGGTACAGATATT 547
Db 308 CGACAAAAAACCTTCTTANTGCAITATCAGAAAGATTTTATAATACAG-GAGGCATATT 366

QY 548 GTCGTTTAAAGAGGGGTCTTAAAGAAAGCACTTGTCTAAGTTAGCAACTGTGAGGATGG 607
Db 367 GCTCAGTCAGAGGGGTCTTATAAGAAAGCACTTACTAAGTTAGCGACTAACAGAACAA 426

QY 608 CCAGTTTAAATAGGACTCAACCCCTCTGGGAGGGACAGCAGGGGGAAGGGGGCT 667
Db 427 CCAGTTTAAAGATGAATTAATGCCCATTTGGGGAGGATGCGCGGTGTTAAGANAAN 486

QY 668 CAAGAGACACTGATAAGATCGGCATTTGTC-ATCTACTGTTTGCAGAAAATTAACCG 726
Db 487 GAAACNTAAGAAAACATTTTCTGTTATACAAACCTTTTNTTNTTATCTACTGNATT 546

QY 727 TTAAGAGCTTTACCGGTGACACTTTTATTCAGTTGAAATTAATCTCCATGTAC 777
Db 547 TGACAAAAATTAACCNCTTTAAAGTTTACCCNGGCACCTNNCTTCTGTC 597

RESULT 13

US-09-871-161-483
; Sequence 483, Application US/09871161
; Publication No. US20030097666A1
; GENERAL INFORMATION:

; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCNDA-260XX
; CURRENT APPLICATION NUMBER: US/09/871.161
; CURRENT FILING DATE: 2001-05-31
; PRIOR APPLICATION NUMBER: 09/328.111
; PRIOR FILING DATE: 1999-06-08
; PRIOR APPLICATION NUMBER: 60/117.393
; PRIOR FILING DATE: 1999-01-27
; PRIOR APPLICATION NUMBER: 60/098.639
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 483
; LENGTH: 605
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(605)
; OTHER INFORMATION: n = A,T,C or G
US-09-871-161-483

Query Match 14.8%; Score 187.2; DB 11; Length 605;
Best Local Similarity 65.5%; Pred. No. 2.9e-40;
Matches 348; Conservative 0; Mismatches 158; Indels 25; Gaps 5;

QY 265 TACCGAGCATGTGTCGGGTGGGTGAGGTCTAGAGGGGGCGGCATCAATCAGATGACA 324
Db 74 TACAGGAGCGGTGTGTCGGGTGGGTGAGGTCTAGAGGGGGCGGCATCAATCAGATGACA 132

QY 325 GTGTTGCTACTCTGGCAAGACAGTGATGTTTCAGAA-----TATCTAA 367
Db 133 GTGTCGTAATCTGGCAAGACAGTGATGTTTCAGAAAGGTTTATAGTTTAAAGAAATTAUCTAA 192

QY 368 AATAGTTTAAAACTGTAAAGCCGACAGCTGATTTCTACCCAGTTACTAGAAAACG 427
Db 193 AATATTTTAAAAACTATAAGCTGCAACACATGATTTTACACCTAGTTACTAGAAAAC 252

QY 428 AAGGAGACACTAGTCAGCTGAGTAAAGGAGGTGAAACAGGAAACGCACTTCTACTATC 487
Db 253 AAGGAAAGCACTTATTAGCTCTGAATA-----AAGTAAACATGGAAGCACTTTTACTAAT 307

QY 488 TACCAAAAAATCTCCGAATGCATTATCAGAAAGATCTTATAGTACAGGTACAGATATT 547
Db 308 CGACAAAAAACCTTCTTANTGCAITATCAGAAAGATTTTATAATACAG-GAGGCATATT 366

QY 548 GTCGTTTAAAGAGGGGTCTTAAAGAAAGCACTTGTCTAAGTTAGCAACTGTGAGGATGG 607
Db 367 GCTCAGTCAGAGGGGTCTTATAAGAAAGCACTTACTAAGTTAGCGACTAACAGAACAA 426

QY 608 CCAGTTTAAATAGGACTCAACCCCTCTGGGAGGGACAGCAGGGGGAAGGGGGCT 667
Db 427 CCAGTTTAAAGATGAATTAATGCCCATTTGGGGAGGATGCGCGGTGTTAAGANAAN 486

QY 668 CAAGAGACACTGATAAGATCGGCATTTGTC-ATCTACTGTTTGCAGAAAATTAACCG 726
Db 487 GAAACNTAAGAAAACATTTTCTGTTATACAAACCTTTTNTTNTTATCTACTGNATT 546

QY 727 TTAAGAGCTTTACCGGTGACACTTTTATTCAGTTGAAATTAATCTCCATGTAC 777
Db 547 TGACAAAAATTAACCNCTTTAAAGTTTACCCNGGCACCTNNCTTCTGTC 597

RESULT 14

US-10-066-543-3022
; Sequence 3022, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun

```

; OTHER INFORMATION: Clone ID: 33-LIB188-020-Q1-B1-A2
US-09-960-352-7612

Query Match          13.6%; Score 172.2; DB 10; Length 432;
Best Local Similarity 72.2%; Pred. No. 2.8e-36;
Matches 294; Conservative 0; Mismatches 103; Indels 10; Gaps 5

QY 379 AACTGTAAGCCGACGACGCTGATTTCTA-CACCAGGTACTAGAAAAAGGAGGAAGCA 437
DB 432 AACCATAAAGCTGCAACACGCTGATTTTATACTTAGTACTAGAAAACTAAGGAAGCA 373
QY 438 CTAGTCAGC--TGAGTAAAGGAAGGTGAAAAACAGGAACGCACCTTCTACTATCTACCAAAA 495
DB 372 CTATTATAGCTTTGAATAAAGCAACATGAAAAACAGGAGTGCACCTTTACTAACTACAAAA 313
QY 496 AATCTCCGAATGATTAACAGAAAGATCTTATAGTACAGTFCAGACATATTGCTCGTTA 555
DB 312 AAAATTTTCTAATGCATATTACAGAAAGATTTTATAATACAAG-GAGGCAATATGCTCATTA 254
QY 556 AGAAGGGGGTCTTAAGAAAAAGCACTTGCTAAAGTTAGCAACTGTGAGGATGCCAGTTTA 615
DB 253 AGAATGAGTCTATAAGAAAGCACTTACTAAGTTAGTACCACTATCAGGATGACCAGGTCA 194
QY 616 AATATGACATCAAGGCC----CCATCTGGGAGGACAGCAGGGGGGAAGGGGGGCTCAAG 671
DB 193 GAGATGATTTAAATGCCCAAATTTTCAGAAGGGGTGGCGCAGTTTAAAGGAAAGCTTAAGAA 134
QY 672 AGAGACACTGATAGATCGGCCATTGTGCATCTACTG--TTTGACAGAAATTAACCGTTA 729
DB 133 ACACTCACTGATATAACCGACCTTCTTCATCCCGCATTTTGACAGAAATTAACCTTTT 74
QY 730 AAAAGCTTTTACCCTGACACTTTTATTTCAGTTGAATTACTCCATGTA 776
DB 73 AAAACTTTTACCCTGATGCTTTTAAATAGTTTAAATGTTTACGTGTA 27

Search completed: November 23, 2003, 19:05:46
Job time : 455.662 secs

```


Db	308	CGACAAAAAACCTTCTAATGCATTATCAGAAAGATTTTATATCAAG-GAGGGCATATT	366
Qy	548	GCTCGTTAAGAAGGGGGTCCTAAGAAAAACACTTGTAAAGTTAGCAACTGTGAGCATGG	607
Db	367	GCTCAGTCAGAAAGGGGTCTATAGAAAAGCACTTACTTAAGTTAGCGACTACAGAACA	426
Qy	608	CCAGTTTAAATATGACTCAACGCGCCCATCTGGGAGGAGCAGACGAGGGGGAAGGGGGCT	667
Db	427	CCNGTTTAAAGATGAATTTAAATGCCCCATTTGGGGANGCATGGCAGTGTTTAAGANAAN	486
Qy	668	CAAGAGAGACACTGATAGATCGGCCATTGTGC-ATCTACTGTTTGACAGAAATTAACCG	726
Db	487	GAAAGCNTPAAGAAACATTTNCTGGTTATANCAAAACCTTNTTNTTATCTACTGNATT	546
Qy	727	TTAAAAAGCTTTTACCCGTGACACTTTTATTCAGTTTGAATTACTCCATGTAC	777
Db	547	TGACAAAAATTAACNTTTTAAAGTTTATCCNGGCACTTNTTCTNTGTGCC	597

```

RESULT 2
US-09-385-982-133
; Sequence 133, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: PRODUCTS: I1
; FILE REFERENCE: CCDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 133
; LENGTH: 588
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(588)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-133

```

Query Match	5.4%; Score 68; DB 3; Length 588;
Best Local Similarity	56.1%; Pred. No. 5.4e-09;
Matches 157; Conservative	0; Mismatches 113; Indels 10; Gaps 4;
Qy	361 TATCTAAAATAGTTTAAAAACGTGAAGAGCCGACAGCTGATTTCTACACCCAGTTACTA 420
Db	189 TATNNANATATNNNNNAANNATTAANCCTGCANCAANTGATTTTNACACCTANTTACTA 248
Qy	421 GAAACGAGGAGGACGACTAGCTGAGTGAAGGAGGTGAAACGAGGAACGCATTC 480
Db	249 GAAACACTAANGAAGGACACTNATTAGCT-----CTGAATNAANTNACATGGAAGCGCTTT 302
Qy	481 TACTATCTACCAAAAAAATCTCCGAATGCATTTATCAGAAAGATCTTTATAGTACGAGTCAG 540
Db	303 TACTAATCTNCAANAANAACCTTC-TCTGCANTATNNNAAGATTTTATNATACA-ANGNG 360
Qy	541 ACATATTCTCTGTTAAGAAGGGGGTCTTAAGAAAGGCATTTGCTAGCTAGCAACTGTG 600
Db	361 GNNNATCNCTNATCATNANGGGTTCATTATTANAAA---CCCTGTAANTNTGCACTTAC 418
Qy	601 AGGATGGCCAGTTTAAATATATGGACTCAACGCCCCCATCTGG 640
Db	419 AGAACANCCAGCNTANANATGANTTTTCATGGCCCAATTGGG 458

RESULT 3
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFELINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-F15
US-08-232-463-14

[illegible]


```

; EARLIER APPLICATION NUMBER: 60/048,949
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,974
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,883
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,897
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,898
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,962
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,963
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,877
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,878
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/070,923
; EARLIER FILING DATE: 1997-12-18
; EARLIER APPLICATION NUMBER: 60/092,921
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/094,657
; EARLIER FILING DATE: 1998-07-30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 103
; LENGTH: 2218
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-205-258-103

Query Match          3.8%; Score 47.8; DB 4; Length 2218;
Best Local Similarity 57.8%; Pred.No. 0.0027;
Matches      85; Conservative    0; Mismatches     62; Indels       0; Gaps      0;

Qy   1 TTTTNTTTTTTTTTTTCAGTCTCCAAAGACATTTTTTTTTTTTTTTTTTGATGTCAAGG 60
Db   2207 TTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATTGTCAACA 2148

Qy   61 ATTATTAACTCATACATGCATAAACAATCTGCTAACTGCATTAGCAAAAAGATCAATGTAA 120
Db   2147 GTCTTTTATTATTATACCTCATAAAGAAGAACAAAGATGATGGTATCAAAGGACAATTTC 2088

Qy   121 AAACACTCCACAATTCGCAACTGTCA 147
Db   2087 AACTAAGAATAGTACATAGCTTTCCA 2061

RESULT 6
US-09-205-258-232/c
; Sequence 232, Application US/09205258
; Patent No. 6525174
; GENERAL INFORMATION:
; APPLICANT: Young et al.
; TITLE OF INVENTION: 207 Human Secreted Proteins
; FILE REFERENCE: PZ007P1
; CURRENT APPLICATION NUMBER: US/09/205,258
; CURRENT FILING DATE: 1998-12-04
; EARLIER APPLICATION NUMBER: PCT/US98/11422
; EARLIER FILING DATE: 1998-06-04
; EARLIER APPLICATION NUMBER: 60/048,895
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,375
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,881
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,880
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,896
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,020
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,877
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,898
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,962
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,917
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,949
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,974
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,883
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,897
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,898
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,962
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,963
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,963
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,877
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,878
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/070,923
; EARLIER FILING DATE: 1997-12-18
; EARLIER APPLICATION NUMBER: 60/092,921
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/094,657
; EARLIER FILING DATE: 1998-07-30
; NUMBER OF SEQ ID NOS: 1227
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 103
; LENGTH: 2218
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-205-258-103

```

```
; NAME/KEY: SITE
; LOCATION: (438)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (741)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-205-258-232

Query Match      3.7%; Score 46.8; DB 4; Length 760;
Best Local Similarity 57.1%; Pred. No. 0.0032;
Matches 84; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

Qy 1 TTTTNTTTTTTTTCAAGTTCCAAAGACATTTTTTTTTTTTTTTTATGATTCGAAG 60
Db 748 TTTTNTTTTTTTTCAAGTTCCAAAGACATTTTTTTTTTTTTTTTATGATTCGAAG 689

Qy 61 ATTATTAAAGTCATACATGCAAAACATCTGTAAGTCAATGCAAAAGATCAATGTAA 120
Db 688 GTGTTTTTATTATACCTACAAAGAAAGAAACAAGATGATGTATCAAAAGGACAAATTTC 629

Qy 121 AAACACTCCACAATTCGCAACTGTCA 147
Db 628 AAACATAGAATAGTAACATAGCTTTCA 602

RESULT 7
US-09-734-675-3/c
; Sequence 3, Application US/09734675
; Patent No. 6365391
; GENERAL INFORMATION:
; APPLICANT: WEBSTER, Marion et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL000862
; CURRENT APPLICATION NUMBER: US/09/734,675
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 38844
; TYPE: DNA
; ORGANISM: Human
; US-09-734-675-3

Query Match      3.6%; Score 45.6; DB 4; Length 38844;
Best Local Similarity 55.8%; Pred. No. 0.036;
Matches 87; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

Qy 2 TTTTNTTTTTTTTCAAGTTCCAAAGACATTTTTTTTTTTTTTTTATGATTCGAAG 61
Db 24608 TCTCTCTATTACTCGAATTCACATAAATCTTTTTTTTTTTTTTTTGGTTAAAGG 24549

Qy 62 TTTATTAAAGTCATACATGCAAAACATCTGTAAGTCAATGCAAAAGATCAATGTAA 121
Db 24548 TTTAGAAATGAAAAACCTTGAAACCTCTTGCATAATGATATTAAATGCTTTTAAAC 24489

Qy 122 AACACTCCACAATTCGCAACTGTCAATGTCAATGTCAATGTCAATGTCAATGTCA 157
Db 24488 ACAGTTCCTAAATAGTAAATAGTCTTTTAAAGAA 24453

RESULT 8
US-09-370-253-1/c
; Sequence 1, Application US/09370253
; Patent No. 6165792
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; APPLICANT: Sakai, Hajime
; APPLICANT: Thorpe, Catherine J.
; TITLE OF INVENTION: Amino Acid Transporters
; FILE REFERENCE: BB-1200
```

```
; CURRENT APPLICATION NUMBER: US/09/370,253
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: 60/097,222
; EARLIER FILING DATE: August 20, 1998
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 1
; LENGTH: 1882
; TYPE: DNA
; ORGANISM: Hordeum vulgare
; US-09-370-253-1

Query Match      3.6%; Score 45; DB 3; Length 1882;
Best Local Similarity 54.5%; Pred. No. 0.015;
Matches 90; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

Qy 1 TTTTNTTTTTTTTCAAGTTCCAAAGACATTTTTTTTTTTTTTTTATGATTCGAAG 60
Db 1860 TTTTNTTTTTTTTCAAGTTCCAAAGACATTTTTTTTTTTTTTTTATGATTCGAAG 1801

Qy 61 ATTATTAAAGTCATACATGCAAAACATCTGTAAGTCAATGCAAAAGATCAATGTAA 120
Db 1800 TTTTNTTTTTTTTCAAGTTCCAAAGACATTTTTTTTTTTTTTTTATGATTCGAAG 1741

Qy 121 AAACACTCCACAATTCGCAACTGTCAATGTCAATGTCAATGTCAATGTCAATGTCA 165
Db 1740 GAACAATCGAATGATGATCTTGGCACTTAATTGATATTCCTC 1696

RESULT 9
US-08-924-747-25/c
; Sequence 25, Application US/08924747
; Patent No. 6063570
; GENERAL INFORMATION:
; APPLICANT: MCGONIGLE, BRIAN
; APPLICANT: O'KEEFE, DANIEL
; TITLE OF INVENTION: SOYBEAN GLUTATHIONE-S-TRANSFERASE
; TITLE OF INVENTION: ENZYMES
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E.I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95
; SOFTWARE: MICROSOFT WORD VERSION 7.0A
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/924,747
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FLOYD, LINDA AXAMEITHY
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: CL-1108
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-8112
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 991 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; TISSUE TYPE: SOYBEAN
```


This Page Blank (uspto)

REFERENCE

AUTHORS

3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsunai, T., Tashiro, H., Itoh, M.,
Suni, I., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

20530913
11076861

4

Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
Fleischmann, M., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G.,
Quackenbush, J., Schriml, L. M., Stauble, F., Suzuki, R., Tomita, M.,
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C.,
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,
Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,
Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,
Toyo-oka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L.,
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S.,
and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

21085660
11217851

5 The PANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

56 (bases 1 to 2269)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurikara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Chisato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Direct Submission

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

Submitted (16-JUL-2001)

Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/
Location/Qualifiers

COMMENT

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

source

1. 2269
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM_DB:9430013G17"
/db_xref="taxon:10090"
/clone="9430013G17"
/tissue_type="embryonic body between diaphragm region and
neck"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="12 days embryo"
198_776
/note="unnamed protein product; RAS-related C3 botulinum
substrate 1 (MGD|GI:97845, GBAK011072, evidence: BLASTN,
100%, match=987)
putative"
/codon_start=1
/protein_id="BAC28767.1"
/db_xref="GI:26330057"
/translation="MQAIKCVVGVGAVGKTCLLSYTTNAPPGEVITPVPNGYANV
MVDGKPNVLGWDVAGQDYLRLPLSYPTQDVLICFLSLYSPSPFENVRKWKYQEV
HCPNTPILVGTUKLDLDDDDKDTIEKLEKRLTPTIYFQGLAWAKIEIGAVKYLECSAL
TQRLGKTVFDEAIRAVLCPFPVKKRKRKLLL"
2252_2257
/note="putative"
2269
/note="putative"
BASE COUNT 555 a 569 c 502 g 643 t
ORIGIN

Query Match

Best Local Similarity

Matches 1081; Conservative

DB 11; Length 2269;

Pred. No. 1.1e-102;

Indels 52; Gaps 10;

QY 50

DB 2269

QY 110

DB 2209

QY 170

DB 2149

QY 230

DB 2089

QY 269

DB 2029

QY 329

DB 1972

QY 389

DB 1912

QY 448

DB 1852

QY 507

DB 1792

QY 563

DB 1733

QY 1675

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 204939374
PubMed 11042159

REFERENCE AUTHORS

3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

TITLE

JOURNAL 20530913
MEDLINE 11076861
PubMed 11076861

REFERENCE AUTHORS

4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Schiavi, R.,
Quackenbush, J., Schriml, L. M., Staib, L. J., Suzuki, R., Tomita, M.,
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C.,
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,
Marchionni, L., Mashima, J., Mazzarelli, J., Mombauts, P., Nordone, P.,
Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schonbach, C., Seiya, T., Shibata, Y., Storch, K. F., Suzuki, H.,
Toyo-oka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L.,
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohsaki, S.,
and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6921), 695-699 (2001)

TITLE

JOURNAL 21085660
MEDLINE 11217851
PubMed 11217851

REFERENCE AUTHORS

5 The PANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.

Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

TITLE

JOURNAL 21085660
MEDLINE 11217851
PubMed 11217851

REFERENCE AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Inotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M., and Hayashizaki, Y.
Direct Submission

TITLE

JOURNAL Submitted (16-JUN-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)

COMMENT

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to

prepare mouse tissues.
Please visit our web site for further details.

URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/.

FEATURES source

1. 2270
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:C130025F01"
/db_xref="taxon:10090"
/clone="C130025F01"
/tissue_type="head"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="16 days embryo"
200..778

CDS

/notes="unnamed protein product; RAS-related C3 botulinum
substrate 1 (MGD|MGI:97845, GB|AK011072, evidence: BLASTN,
100%, match=987)
putative"

/codon_start=1
/protein_id="BAC33203.1"
/db_xref="GI:26339064"
/translation="MQAIKVVGDAVGKTCLLISYTTNAPGEVPTVFDNYSNV
VMDKRPVNLGMDTAGQEDYDLRLPLSYFQTDVFLICFLSVSPASENVRKRYPEVR
HHCPTPIILVGTGKLDKDDKDTIEKLEKLLTPITYQGLAMAKBIGAVKYLECSAL
TORGKTVDFEAIKRAVLCPPVKKRKKCLLL"

2254..2259
/note="putative"

2270
/note="putative"

BASE COUNT 554 a 569 c 503 g 644 t

ORIGIN

Query Match 67.1%; Score 849.6; DB 11; Length 2270;

Best Local Similarity 85.6%; Pred. No. 2.1e-102;

Matches 1079; Conservative 0; Mismatches 129; Indels 52; Gaps 10;

QY 52 GATTCAAGGATTATTAAAGTCATACATGCAAAACATACCTGCTAAGTCTAGTCAATAGCAAAAGA 111

DB 2269 GATTCAAGGATTATTAAATTCATACATGCAAAACATACCTGCTAAGTCTAGTCAATAGCAAAAGA 2210

QY 112 TCATGTAAACACATCCCAANTCTGCAACTGTCAATGAAAGGTTTGTCTAGTGG 171

DB 2209 TCAATGTAAACACATCCCAANTCTGCAACTGTCAATGAAAGGTTTGTCTAGTGG 2150

QY 172 TCGAAGGCGCCCAACTGCTGTTCTTCCAGTGTAGTGTGTACAGACGGCGTTAGCA 231

DB 2149 TTGAAGGCGCCCAACTGCTGTTCTTCCAGTGTAGTGTGTACAGACGGCGTTAGCA 2090

QY 232 CTAGCGCTTGACAGAACCTCCTACAGACCCCAAGG-----TACC GG 270

DB 2089 CTAGCGCTTGACAGAACCTCCTACAGACCCCAAGG-----TACC GG 2030

QY 271 AGCATGTCTCGCGTGGTGGTCTAGAGGGGGCGGCATCAATCAATGACAGTGTG 330

DB 2029 AAGCGCTCTTCCACGTGGCGGAGGCTAGA--GGCAGCATTAGTACATGACAGTGTG 1973

QY 331 GTACTCTGCAAGACAGTGTGTTTTCAGAAATCTAAATAGTTTAAACTGTAAAGCC 390

DB 1972 GTACTCTGCAAGGTTTCATAGTTTAAAGAAATCTAAATAGTTTAAACCTGAAAGCT 1913

QY 391 GCAGCAGCTGATTCTTACACCCAGTCTACAGAAACGAGGGAAGCACTAGT-CAGCTCA 449

DB 1912 GCACACATGATTTCACACCTAGTTTGTAGAAACTAGGAAAGCACTAGTCTGTA 1853

QY 450 GTAAGGAGGTGAAACAGGAACCCACTTCTATCTATCTACCAAAAAAATCTCCGAA--- 506

DB 1852 GTAAAGTAGGCGAAACAGGACGCACTTCTACTGAGCTACCAAAAAAATAAATTC 1793

QY 507 --TGCAATTATCAGAAAGATCTTATAGTACAGTCTAGATATGCTCTGTAGAGGGGG 564

DB 1792 TCTGCGTTATCAGAAAGATCTTATAGTACAGGTCAGATATGCTCTGTAGAGGGGG 1734

```

565 TCCTAAAGAAAGACCTTGCTAAGCTTACGACTGTGAGATGGCGCAGTTTAAATATGAC 624
Db TTCTAAAGAAAGACCTTGCTAAGCTTACGACTGTGAGATGGCGCAGTTTAAATATGAC 1675

625 TCAACGCCCATCTGGGGAGGACAGCAGGAGGGGAGGGGGCTCAAG-----A 672
Db TCAACGCCCATCTGGGGAGGACAGCAGTGGAGGGTGGGGGCGCGAAGGATGCTCAA 1615

673 GAGACACTGATAAGATCGGCATTTGTCACTTACTGTTTGTGACAGAAATTAACCGTTAAAA 732
Db GAGACACTGATAAGATCGGCATTTGTCACTTACTGTTTGTGACAGAAATTAACCGTTAAAA 1555

733 AGCTTTACCGCTGACACTTTTATTCAGTTGAAATTAATCTCCATGTAACAATGTAGTGAATTT 792
Db AGCTTTACCGCTGACACTTTTATTCAGTTGAAATTAATCTCCATGTAACAATGTAGTGAATTT 1495

793 AATCTCTACTTCAATATTAGTCAAAATACCTGTCTCTCTCTTTGTAGTGCCTGTGTTTCCAC 852
Db AATCTCTACTTCAATATTAGTCAAAATACCTGTCTCTCTCTTTGTAGTGCCTGTGTTT--C 1437

853 ACATCTCCACCCAGCACCACCCAGCTAGGACAGAAATCTTCGTTAGAGGCAACACAGGA 912
Db ACATCTCCACCCAGCACCACCCAGCTAGGACAGAAATCTTCATTTAGAGGCAACACAGGA 1377

913 GCCAGAGTCTGTTCAAAAGCCTGCGAAGCCGCTGAGCTGCTGTTATTTTAGAGAACTCACTA 972
Db ACCAGAGTCTGTTCAAAAGTCTGCGAAGCTGAGCTGCTGTTATTTTAGAGAACTCACTA 1317

973 TGAATCAAAAGAGCAGAGCTGTTACACCCCATC- --GTGACGTACAGTACAAAGTTACGTA 1029
Db TGAATCAAAAGAGCTGAGCTGTCACACTCATCTGTGACGTACAGTACAAAGTTACGTA 1257

1030 ATGACATGGGCTGATAGTTACAGTGGGTTACATGGCAGCGTGTCTATTAGAGGAGCTG 1089
Db ATGACATGGGCTGATAGTTACAGTGGGTTACATGGCAGCGTGTCTATTAGAGGAGCTG 1197

1090 TGCTGTGTCACAGCTGCTGGAGCTACGGAGGGTCTGACCCCTTGAGCCCAAGAGCTGC 1149
Db TGCTGTGTCACAGCTGCTGGAGCTACGGAGGGTCTGACCCCTTGAGCCCAAGAGCTGC 1137

1150 AGTCTTCTTAAAGACAAA- --GTCTCTCAACAGCTAGTGTCTTACGTGTTCTCAGCACAA 1206
Db GGTCTTCTTAAAGACAGAGTCTCTCAAGAGAGTTAGTGTCTTACGTGTTCTCAGCACAA 1077

1207 CGCACTTAGTTCACAGTATTTTGGCAATTTCTTAATCTGAGCAAGATAGGGATTTT 1266
Db CGCACTTAGTTCACAGGATTTTGGCAATTTTAAATCTGAGCAAGATAGGGGATTTT 1017

RESULT 4
AK076023/c
LOCUS 2308 bp mRNA linear HTC 07-DEC-2002
DEFINITION Mus musculus 10 days embryo whole body cDNA, RIKEN full-length
subtracted library, clone:2610100016 product:RAS-related C3 botulinum
substrate 1, full insert sequence.
ACCESSION AK076023
VERSION AK076023.1 GI:26344957
KEYWORDS HTC; cap-trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
TITLE Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
TITLE Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
JOURNAL Normalization and subtraction of cap-trapper-selected cDNAs to
prepare mouse tissues.

```

```

JOURNAL MEDLINE PUBMED REFERENCE
AUTHORS
3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuura, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yanamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multiplexed capillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861
4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
Fleischmann, M., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G.,
Quackenbush, J., Schriml, L. M., Staubli, F., Suzuki, R., Tomita, M.,
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
Baidarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Hill, D.,
Flatcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,
Machionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,
Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,
Toyo-Oka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L.,
Wyshak-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kotsuki, S.
and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
21085660
11217851
5 The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 2308)
Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,
Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y.,
Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K.,
Hiraoka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,
Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M.,
Koya, S., Kurahara, C., Matsuyama, T., Miyazaki, A., Nishi, K.,
Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C.,
Saito, H., Saito, R., Sakai, K., Sakai, K., Sano, H., Sasaki, D.,
Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y.,
Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T.,
Tejima, Y., Toya, T., Yamamura, T., Yamanaka, I., Yasunishi, A.,
Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-resgsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.

```

COMMENT

Please visit our web site for further details.

URL: <http://genome.gsc.riken.go.jp/>
URL: <http://fantom.gsc.riken.go.jp/>

FEATURES

source	Location/Qualifiers
1. .2308	/organism="Mus musculus"
	/mol_type="mRNA"
	/strain="C57BL/6J"
	/db_xref="FANTOM DB:2610100016"
	/db_xref="MGI:1898616"
	/db_xref="taxon:10090"
	/clone="2610100016"
	/tissue type="whole body"
	/clone lib="RIKEN full-length enriched mouse cDNA library"
	/dev_stage="10 days embryo"
239. .817	
	/notes="unnamed protein product; RAS-related C3 botulinum substrate 1 (MG MG1:97845, GB AK011072, evidence: BLASTN, 100%, match=987)"
	putative"
	/codon_start=1
	/protein_id="BAC36128.1"
	/db_xref="GI:26344958"
	/db_xref="MGI:97845"
	/translation="MQAIKVVGVGDGAVGKTCLLISYITNAPGEVPIPTVFONYSANV MVDGKPNVLGMDTAGEDYDRPLSPYDPTDFELICFSLVSAASPENVRKYPYR HHCNPPIILVGTPLKLDKDDKDFIEKLKELKLPYIPYPOGLAWAKELGAVKYLECSAL TORGLKTVFDEAIRAVLCPPPVKKRKKCLLL"
	2293. .2298
	/notes="putative"
	2308
	/notes="putative"
BASE COUNT	559 a 580 c 518 g 650 t 1 others
ORIGIN	
Query Match	64.4%; Score 815.4; DB 11; Length 2308;
Best Local Similarity	84.9%; Pred. No. 6.1e-98;
Matches 1071; Conservative	0; Mismatches 136; Indels 54; Gaps 12;
QY	52 GATTCAGGATTTATTAAAGTATACATGCAAAACATACCTGCTAACTGCAATGCAAAAAGA 111
DB	2308 GATTCAGGATTTATTAAATTCATACATGCAAAACATACCTGCTAACTGCAATGCAAAAAGA 2249
QY	112 TCAATGTAAACACATCCCAATTCGCACTGCTCAATTCGAAAGAGTTTGTCTAGTGG 171
DB	2248 TCGATGTAAACACATCCCAATTCGCACTGCTCAATTCGAAAGAGTTTGTCTAGTGG 2189
QY	172 TCGAAGGCCCAACACTGTGTTCTGCCAGTGAAGTTAGTTGTACAGAACGGCGGTAGCA 231
DB	2188 TTGAGAGGCCCAACACTGTCATCTTCCAGTGAAGTTAGTTGTACAGAACGGCGGTAGCA 2130
QY	232 CTAGCGCTTGACAGACCTCAGACCCCAAGG-----TACCGG 270
DB	2129 CTAGCAGTTGACAGACCTCAGACCCCAAGGACATCTCTAGGCAGACCCATGACAGG 2070
QY	271 AAGCATGTCTCCGCTGGGTGAGGTCTAGAGGGGGGCGATCAATCACATGACAGTGTG 330
DB	2069 GAGCGTCTCCAGTGGCGAGGCTTAGA---GGCAGCAATAGTACATGACAGTGTG 2013
QY	331 GTACTCTGGCAAGACAGTGAATGTTTCAGAAATCTCTAAATAGTTTAAAACTGTAAAGCC 390
DB	2012 GTACTGTTTAGAGGTTTCATAGTTTAAAGATATCTAAATAGTTTAAAACTGTAAAGT 1953
QY	391 GCAGCAGCTGATTTCTACACCCAGTACTAGAAACGAGAGGAGCACTAGT-CAGCTGA 449
DB	1952 GCAACACATGATTTGCACACCTAGTTGCTAGGAACTAAGGAAGCACTAAGTCTCTGA 1893
QY	450 GTAAAGGAGGTGAACACAGGAACGCACTCTACTATCTACCAAAAATCTCCGAA---506
DB	1892 GTAAAGTAAAGCGAACAAGGACGCACTCTACTGAGCTACCAAAAAAACAATAATC 1833
QY	507 --TGCAATTATCAGAAGATCTTATAGTACAGGTCAGACATATGCTGTTAAGAGGGGG 564

Db	1832 TCTCGGTTATCAGAAAGATCTTATAGTACAGGTGAGACATATTCCTGTTAAGAA-GGGG 1774
QY	565 TCCTAAAGAAAAGCAGCTTGTCTAGTAACTGTGAGGATGGCCAGTTTAAATATGAC 624
DB	1773 TTCTAAAGAAAAGCAGCTTGTCTAGTAACTGTGAGGATGGCCAGTTTAAATATGAC 1715
QY	625 TCAACGCCCATCTCTGGGAGGAGCAGCAGGGGGAAGGGGGCTCAAG-----A 672
DB	1714 TCAACGCCCATCTCTGGGAGGAGCAGCAGTGGAGGTTGGGGGGCGGAGAGGATGCTCAA 1655
QY	673 GAGACACTGATAAGATCGCCATTTGCTACTACTGTTTACAGAAATTAACCGTTAAAA 732
DB	1654 GAGACACTGATAAGATCGCCATTTGCTACTACTGTTTACAGAAATTAACCGTTAAAA 1595
QY	733 AGCTTTACCGGTGACACTTTTATTCAAGTTGAATTAATCTCAATGTACAAATGTAGTGAAT 792
DB	1594 AGCTTTACCGGTGACACTTTTATTCAAGTTGAATTAATCTCAATGTACAAATGTAGTGAAT 1535
QY	793 AATCTCTACTTCAATATTAGTCAAAATCTGCTGTCTCTCTTTGATGACGTCGTGTTTAC 852
DB	1534 AATCTCTACTTCAATATTAGTCAAAATCTGCTGTCTCTCTTTGATGACGTCGTGTTTAC 1477
QY	853 ACACCTCCACCCAGCAGACCCACCTAGGAAACAGATACCTTCTGTTAGAGCAACACAGGA 912
DB	1476 ACACCTCCACCCAGCAGACCCACCTAGGAAACAGATACCTTCTGTTAGAGCAACACAGGA 1417
QY	913 GCAGAGTTTCTGTCTCAAGCCTGCAAGCGGCTCAGCTGCTGTTTATTTAGAGAACTCACTA 972
DB	1416 ACCAGAGTTTCTGTCTCAAGCCTGCAAGCGGCTCAGCTGCTGTTTATTTAGAGAACTCACTA 1357
QY	973 TGAATCAAGAGCAGAGCTGTTACACCATC---GTGAGCTACAGTACAAAGTTAGCTA 1029
DB	1356 TGAATCAAGAGCAGCTGAGCTGTACACTCATCTGTCAGTACAGTACAAAGTTAGCTA 1297
QY	1030 ATGAGCATGGCTGATAAGTTACAGTGTGTTACATGGCAGCGTGTCAATTAAGAGGCTG 1089
DB	1296 ATGACATGGCTGATAGTTTACAGTGTGTTACATGGCAACGTGTCAATTAAGAGGCTG 1237
QY	1090 TGTGTGTACACGGTCTCTGGAGCTACGGGAGGCTCTGCACCCCT-GAGCCCAAGAGCTG 1148
DB	1236 TGTGTGTACACGGTCTCTGGAGCTACGGGAGGCTCTGCACCCCTGCTTCCAGAGCTG 1177
QY	1149 CAGTCTCTTAGGACAAA---GTCTCTCAACAGCTTAGTGTCTGTTCTCAGACA 1205
DB	1176 CGGTCTCTTAGCAACAGAAAGTCTCTCAAGAGAGTTAGTGTCTGTTCTCAGACA 1117
QY	1206 AGCAACTTAGTTCACAGGTATTTGGCAATTTCTAATCTGAGCAAGATAGGGGATTT 1265
DB	1116 AGCCAACTAGTTCAGAGGTATTTGGCAATTTTAAATCTGAGCAAGATAGGGGATTT 1057
QY	1266 T 1266
DB	1056 T 1056
RESULT 5	
BQ191985	
LOCUS	BQ191985 760 bp mRNA linear EST 30-APR-2002
DEFINITION	UI-R-DR1-cky-h-11-0-UI-s1 UI-R-DR1 Rattus norvegicus cDNA clone
ACCESSION	UI-R-DR1-cky-h-11-0-UI 3', mRNA sequence.
VERSION	BQ191985
KEYWORDS	BQ191985.1 GI:20367536
SOURCE	EST.
ORGANISM	Rattus norvegicus (Norway rat)
REFERENCE	1 (bases 1 to 760)
AUTHORS	Ronald, M.F., Lennon, G. and Soares, M.B.
TITLE	Normalization and subtraction: two approaches to facilitate gene
JOURNAL	discovery
	Genome Res. 6 (9), 791-806 (1996)


```

QY 573 AAAAGCACTGCTAAGTTAGCAACTGTGAGGATGGCCAGTTTAAATATGACTCAAGGCC 632
    |||||
Db 542 AAAAGCACTGCTAAGTTAGCAACTGTGAGGATGGCCAGTTTAAATATGACTCAAGGCC 601
    |||||
QY 633 CCACTCTGGGAGGACAGCAGGGGGAAGGGGGCTCAAG 671
    |||||
Db 602 CCACTCTGGGAGGACAGCAGGGGGAAGGGGGCTCAAG 640
    |||||

RESULT 9
CA339407/c
LOCUS
DEFINITION
NISC ly02d09.y1 NCI CGAP Pr32 Rattus norvegicus cDNA clone
IMAGE:5622185 5', mRNA sequence.
ACCESSION
CA339407.1 GI:24557505
VERSION
EST.
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 644)
REFERENCE
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
cDNA Library Preparation:
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC)
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
info@image.llnl.gov
Plate: LLAM12447 row: G column: 18
Seq primer: M13RPI reverse primer (ABT).
FEATURES
Location/Qualifiers
1..644
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="IMAGE:5622185"
/sex="male"
/tissue_type="dorsal prostate"
/dev_stage="adult, 14 month"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI CGAP Pr32"
/note="Organ: prostate; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2 kb. Constructed by Invitrogen. Note:
this is a NCI-CGAP Library."
BASE COUNT 145 a 164 c 129 g 206 t
ORIGIN

Query Match 49.9%; Score 631.4; DB 14; Length 644;
Best Local Similarity 99.8%; Pred. No. 1.4e-73;
Matches 632; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 113 CAATGTAAACCACTCCCAATTCGCACTGCAATTCGAAATGTTTCTAGTGGT 172
    |||||
Db 644 CAATGTAAACCACTCCCAATTCGCACTGCAATTCGAAATGTTTCTAGTGGT 585
    |||||
QY 173 CGAAGGCCCACTGTGTTCTTGCCAGTGTAGTTGTACAGAACGGCGTTAGCAC 232
    |||||
Db 584 CGAAGGCCCACTGTGTTCTTGCCAGTGTAGTTGTACAGAACGGCGTTAGCAC 525
    |||||
QY 233 TAGCGCTTGACAGACCTTCACAGACCCCAAGTACCGAAGCATGTGTCGGTGGTGA 292
    |||||
Db 524 TAGCGCTTGACAGACCTTCACAGACCCCAAGTACCGAAGCATGTGTCGGTGGTGA 465
    |||||
QY 293 GGTCTAGAGGGGGCGCATCAATCATGACAGTGTGTGTTCTTGGCAGACAGTGATG 352
    |||||

```

```

Db 464 GGTCTAGAGGGGGCGCATCAATCATGACAGTGTGGTACTCTGGCAACAGCAGTGATG 405
    |||||
QY 353 TTTTCAGAAATATCTAAATAAGTTTAAACCTGTAAAGCCGAGCAGCGTGAATTTCTACACCC 412
    |||||
Db 404 TTTTCAGAAATATCTAAATAAGTTTAAACCTGTAAAGCCGAGCAGCGTGAATTTCTACACCC 345
    |||||
QY 413 AGTTACTAGAAAAACGAAGGAAGCAGTGTAGTCAGTGTAGTAAAGAGAGGTGAAAAACAGGAA 472
    |||||
Db 344 AGTTACTAGAAAAACGAAGGAAGCAGTGTAGTCAGTGTAGTAAAGAGAGGTGAAAAACAGGAA 285
    |||||
QY 473 CGCAGCTTCTACTATCTACCAAAAAAATCTCCGAATGCATTATCAGAAGATCTTTATAGTA 532
    |||||
Db 284 CGCAGCTTCTACTATCTACCAAAAAAATCTCCGAATGCATTATCAGAAGATCTTTATAGTA 225
    |||||
QY 533 CAGGTCAGACATATTCGTTAAGAGAGGGGTCCTAAAGAGAGAGCCTTGTCTAAGTTAG 592
    |||||
Db 224 CAGGTCAGACATATTCGTTAAGAGAGGGGTCCTAAAGAGAGAGCCTTGTCTAAGTTAG 165
    |||||
QY 593 CAACCTGTAGGATGGCCAGTTTAAATATGAGACTCAACGCCCCCATCTGGGAGGAGCAGCA 652
    |||||
Db 164 CAACCTGTAGGATGGCCAGTTTAAATATGAGACTCAACGCCCCCATCTGGGAGGAGCAGCA 105
    |||||
QY 653 GGGGGAAGGGGGCTCAAGAGACACACTGATAGATCGGCCATTGTCTACTCTGTTG 712
    |||||
Db 104 GGGGGAAGGGGGCTCAAGAGACACACTGATAGATCGGCCATTGTCTACTCTGTTG 45
    |||||
QY 713 ACAGAAATTAACCGTTAAAGAGCCTTTACCCGTG 745
    |||||
Db 44 ACAGAAATTAACCGTTAAAGAGCCTTTACCCGTG 12
    |||||

RESULT 10
BI283790
LOCUS
DEFINITION
UI-R-CW08-cch-g-03-0-UI.s1 UI-R-CW08 Rattus norvegicus cDNA clone
UI-R-CW08-cch-g-03-0-UI 3', mRNA sequence.
ACCESSION
BI283790
VERSION
EST.
KEYWORDS
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 638)
REFERENCE
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
The sequence contained an oligo-dt track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dt track served to identify it as a clone from the
non-normalized rat aorta pool library cDNA Library Preparation:
M.B. Soares Lab Clone distribution: clones will be available
through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.
Location/Qualifiers
1..638
/organism="Rattus norvegicus"
/mol_type="mRNA"

```

```
/strain="Sprague-Dawley"
/db xref="taxon:10116"
/clone="UI-R-CW08-cch-g-03-0-UI"
/dev stage="ADULT"
/lab host="DHI0B (Life Technologies)"
/clone lib="UI-R-CW08"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not 1; Site 2: Eco RI; The UI-R-CW08
library is a non-normalized library constructed from the
following rat aorta tissues: embryonic day 19, embryonic
day 21, adult day 1, adult day 12, adult day 15, adult day
200. For a detailed description of the library from which
this clone was derived, please visit our web site at
ratest.eng.uiowa.edu. The subtraction has been previously
described in (Bonaldo, Lennon and Soares, Genome Research
6:791-806, 1996)
TAG LIB=UI-R-CW08
TAG TISSUE=rat aorta pool
TAG_SEQ=CTGTAGATC"
BASE COUNT 210 a 121 c 149 g 158 t
ORIGIN
Query Match 49.6%; Score 627.4; DB 12; Length 638;
Best Local Similarity 99.1%; Pred. No. 4.8e-73;
Matches 631; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 33 TTTT...TATGATTCAGGATTTATTAAGTCATACATGCAAAACATACCTGC 92
Db 2 TTTT...TATGATTCAGGATTTATTAAGTCATACATGCAAAACATACCTGC 61
QY 93 TAACTGCATTAGCAAAAGATCAATGTAAACACCTCCCAATTCCTGCAACTGTCAATTGA 152
Db 62 TAACTGCATTAGCAAAAGATCAATGTAAACACCTCCCAATTCCTGCAACTGTCAATTGA 121
QY 153 AAAAGTTTGTCTAGTGGTGAAGGCCCAACACATGTGTTCTTCCAGTGAGTTAGTT 212
Db 122 AAAAGTTTGTCTAGTGGTGAAGGCCCAACACATGTGTTCTTCCAGTGAGTTAGTT 181
QY 213 GTACAGACGGGTAGCACTAGCCCTTCACAGACCTTCACAGACCTTCACAGACCTTCACAGAC 272
Db 182 GTACAGACGGGTAGCACTAGCCCTTCACAGACCTTCACAGACCTTCACAGACCTTCACAGAC 241
QY 273 GCATGTGTCGGGTGAGTGTCTAGAGGGGGGGGGCATCAATCATCATGACAGTGTGGT 332
Db 242 GCATGTGTCGGGTGAGTGTCTAGAGGGGGGGGGCATCAATCATCATGACAGTGTGGT 301
QY 333 ACTCTGGCAAGACAGTGTATTTTCAAGATATCTAAATAGTTTAAACCTGTAAGCGGC 392
Db 302 ACTCTGGCAAGACAGTGTATTTTCAAGATATCTAAATAGTTTAAACCTGTAAGCGGC 361
QY 393 AGCAGTGTATTTCTACCCAGTGTACTAGAAACGAGGAGGAGCTAGTCAAGTGA 452
Db 362 AGCAGTGTATTTCTACCCAGTGTACTAGAAACGAGGAGGAGCTAGTCAAGTGA 421
QY 453 AAGGAAGGTGAAACAGGACGACCTTCTACTATCTACCAAAAAATCTCCGAATGCATT 512
Db 422 AAGGAAGGTGAAACAGGACGACCTTCTACTATCTACCAAAAAATCTCCGAATGCATT 481
QY 513 ATCAGAAAGATCTTATAGTACAGGTGAGTGTCTGTTTAAAGAGGGGGTCTTAAG 572
Db 482 ATCAGAAAGATCTTATAGTACAGGTGAGTGTCTGTTTAAAGAGGGGGTCTTAAG 541
QY 573 AAAAGCACTTGTAGTTAGCACTGTAGAGGTGGCAGTTTAAATATGACTCAAGCGC 632
Db 542 AAAAGCACTTGTAGTTAGCACTGTAGAGGTGGCAGTTTAAATATGACTCAAGCGC 601
QY 633 CCATCTGGGAGGGAGCAGCAGGGGGAGGGGGGCTCA 669
Db 602 CCATCTGGGAGGGAGCAGCAGGGGGAGGGGGGCTCA 638
RESULT 11
BE111691
```

```
LOCUS BE111691 639 bp mRNA linear EST 13-JUN-2000
DEFINITION UI-R-BJ1-avv-f-10-0-UI.s1 UI-R-BJ1 Rattus norvegicus cDNA clone
UI-R-BJ1-avv-f-10-0-UI 3', mRNA sequence.
ACCESSION BE111691
VERSION BE111691
KEYWORDS EST.
SOURCE BE111691.1 GI:8503796
ORGANISM Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 639)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized ventricle at 13 dpc library cDNA Library Preparation:
M.B. Soares Lab Clone distribution: clones will be available
through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.
Location/Qualifiers
1..639
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BJ1-avv-f-10-0-UI"
/clone lib="UI-R-BJ1"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not 1; Site 2: Eco RI; The UI-R-BJ1
library is a subtracted library derived from the following
tissues: atrium at 16.5 dpc, ventricle at 16.5 dpc, AV
canal at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc,
AV canal at 15 dpc, ventricle at 13 dpc, and adult heart.
For a detailed description of the library from which this
clone was derived, please visit our web site at
ratest.eng.uiowa.edu. The subtraction has been previously
described in (Bonaldo, Lennon and Soares, Genome Research
6:791-806, 1996)
TAG LIB=UI-R-BJ1
TAG TISSUE=ventricle at 13 dpc
TAG_SEQ=CAGCGA"
BASE COUNT 209 a 121 c 151 g 157 t 1 others
ORIGIN
Query Match 49.3%; Score 624.6; DB 10; Length 639;
Best Local Similarity 99.2%; Pred. No. 1.1e-72;
Matches 627; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 33 TTTT...TTTATGATTCAGGATTTATTAAGTCATACATGCAAAACATACCTGC 92
Db 1 TTTT...TTTATGATTCAGGATTTATTAAGTCATACATGCAAAACATACCTGC 60
QY 93 TAACTGCATTAGCAAAAGATCAATGTAAACACCTCCCAATTCCTGCAACTGTCAATTGA 152
Db 61 TAACTGCATTAGCAAAAGATCAATGTAAACACCTCCCAATTCCTGCAACTGTCAATTGA 120
```

```
QY 153 AAAAGTTTGTCTAGTGGTCCAAAGGCCCAACACTGTGTTCTTGCCAGTGGTTAGTT 212
Db 121 AAAAAGTTTGTCTAGTGGTCCAAAGGCCCAACACTGTGTTCTTGCCAGTGGTTAGTT 180
QY 213 GTACAGAACGGCGTTAGCACTAGCGCTTGA CAGAACCTTCACAGACCCCAAGGTTACCGGAA 272
Db 181 GTACAGAACGGCGTTAGCACTAGCGCTTGA CAGAACCTTCACAGACCCCAAGGTTACCGGAA 240
QY 273 GCATGTGTCCGCGGGGTGAGTCTAGAGGGGGGGGCGCATCAATCACATGACAGTGTGGT 332
Db 241 GCATGTGTCCGCGGGGTGAGTCTAGAGGGGGGGGCGCATCAATCACATGACAGTGTGGT 300
QY 333 ACTCTGCAAGACAGTGTATGTTTCAGAAATATCTAAATAGTTTAAAACTGTTAAAGCGC 392
Db 301 ACTCTGCAAGACAGTGTATGTTTCAGAAATATCTAAATAGTTTAAAACTGTTAAAGCGC 360
QY 393 AGCAGTGTATTTACACCCAGTTACTAGAAAAAGGAGGAGCACTAGTCAAGTGAAGTA 452
Db 361 AGCAGTGTATTTACACCCAGTTACTAGAAAAAGGAGGAGCACTAGTCAAGTGAAGTA 420
QY 453 AAGGAAGGTGAACACAGGACGCACTTCTACTATCTACCAAAAAAATCTCCGAATGCATT 512
Db 421 AAGGAAGGTGAACACAGGACGCACTTCTACTATCTACCAAAAAAATCTCCGAATGCATT 480
QY 513 ATCAGAAAGATCTTATAGTACAGTCAAGTCAACATATGCTGTAAAGGGGGTCTTAAAG 572
Db 481 ATCAGAAAGATCTTATAGTACAGTCAAGTCAACATATGCTGTAAAGGGGGTCTTAAAG 540
QY 573 AAAAGCACTTGTAAAGTACCACTGTGAGATGGCCAGTTTAAATATGACTCAAGCC 632
Db 541 AAAAGCACTTGTAAAGTACCACTGTGAGATGGCCAGTTTAAATATGACTCAAGCC 600
QY 633 CCATCTGGGAGGACAGAGCGGGGGAGGGG 664
Db 601 CCATCTGGGAGGACAGAGCGGGGGAGGGGG 632

RESULT 12
BQ780658
LOCUS
DEFINITION
UI-R-PF0-cpc-22-0-UI.s1 UI-R-PF0 Rattus norvegicus cDNA clone
BQ780658
VERSION
BQ780658.1 GI:21989130
KEYWORDS
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1. (bases 1 to 636)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL
MEDLINE
97044477
PUBMED
8889548
COMMENT
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Jeff Stevens
cDNA Library prepared by: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: DISTRIBUTION: Researchers may obtain clones
from Research Genetics (www.resgen.com).
Seq primer: M13 FORWARD
POLYA=Yes.
Location/Qualifiers
```

```
source
1. .636
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="UI-R-PF0-cpc-22-0-UI"
/tissue_type="Mixed tissues"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-R-PF0"
/note="Vector: p77T3-Pac (Pharmacia) with a modified
polylinker; Site 1: EcoR I; Site 2: Not I; UI-R-PF0 is a
subtracted cDNA library containing the following tissue(s)
): Normal cartilage and SR-JWS Tumor Line . The
subtraction was made according to Bonaldo, Lennon and
Soares, Genome Research, 6:791-806, 1996. The
oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dr)18 tail. The
sequence tags for these libraries are: CTAATGGAGC,
CATCTTTGTA.
TAG LIB=UI-R-PF0
TAG_TISSUE=cartilage
TAG_SEQ=CTAATGGAGC"
BASE COUNT 210 a 120 c 148 g 158 t
ORIGIN
Query Match 49.3%; Score 623.8; DB 13; Length 636;
Best Local Similarity 98.9%; Pred. No. 1.4e-72;
Matches 628; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 33 TTTTCTTTTCTTTTATGATTCAGGATTTTAAAGTCATACATGCAAAACATCTGC 92
Db 2 TTTTCTTTTCTTTTATGATTCAGGATTTTAAAGTCATACATGCAAAACATCTGC 61
QY 93 TAACTGCATTAGCAAAAGATCAATGTAAACACTCCAAATTCGCAACTGCAATTGA 152
Db 62 TAACTGCATTAGCAAAAGATCAATGTAAACACTCCAAATTCGCAACTGCAATTGA 121
QY 153 AAAAAGTTTGTCTAGTGGTCCAAAGGCCCAACACTGTGTTCTTGCCAGTGGTTAGTT 212
Db 122 AAAAAGTTTGTCTAGTGGTCCAAAGGCCCAACACTGTGTTCTTGCCAGTGGTTAGTT 181
QY 213 GTACAGAACGGCGTTAGCACTAGCGCTTGA CAGAACCTTCACAGACCCCAAGGTTACCGGAA 272
Db 182 GTACAGAACGGCGTTAGCACTAGCGCTTGA CAGAACCTTCACAGACCCCAAGGTTACCGGAA 241
QY 273 GCATGTGTCCGCGGGGTGAGTCTAGAGGGGGGGGCGCATCAATCACATGACAGTGTGGT 332
Db 242 GCATGTGTCCGCGGGGTGAGTCTAGAGGGGGGGGCGCATCAATCACATGACAGTGTGGT 301
QY 333 ACTCTGCAAGACAGTGTATGTTTCAGAAATATCTAAATAGTTTAAAACTGTTAAAGCGC 392
Db 302 ACTCTGCAAGACAGTGTATGTTTCAGAAATATCTAAATAGTTTAAAACTGTTAAAGCGC 361
QY 393 AGCAGTGTATTTACACCCAGTTACTAGAAAAAGGAGGAGCACTAGTCAAGTGAAGTA 452
Db 362 AGCAGTGTATTTACACCCAGTTACTAGAAAAAGGAGGAGCACTAGTCAAGTGAAGTA 421
QY 453 AAGGAAGGTGAACACAGGACGCACTTCTACTATCTACCAAAAAAATCTCCGAATGCATT 512
Db 422 AAGGAAGGTGAACACAGGACGCACTTCTACTATCTACCAAAAAAATCTCCGAATGCATT 481
QY 513 ATCAGAAAGATCTTATAGTACAGTCAAGTCAACATATGCTGTAAAGGGGGTCTTAAAG 572
Db 482 ATCAGAAAGATCTTATAGTACAGTCAAGTCAACATATGCTGTAAAGGGGGTCTTAAAG 541
QY 573 AAAAGCACTTGTAAAGTACCACTGTGAGATGGCCAGTTTAAATATGACTCAAGCC 632
Db 542 AAAAGCACTTGTAAAGTACCACTGTGAGATGGCCAGTTTAAATATGACTCAAGCC 601
QY 633 CCATCTGGGAGGACAGAGCGGGGGAGGGGGCT 667
Db 602 CCATCTGGGAGGACAGAGCGGGGGAGGGGGCT 636
```



```

RESULT 13
BG380448      615 bp      mRNA      linear      EST 12-MAR-2001
LOCUS        UI-R-CT0-btx-d-05-0-UI-s1 UI-R-CT0 Rattus norvegicus cDNA clone
DEFINITION   UI-R-CT0-btx-d-05-0-UI 3', mRNA sequence.
ACCESSION   BG380448
VERSION     BG380448.1 GI:13304920
KEYWORDS    Rattus norvegicus (Norway rat)
SOURCE      Rattus norvegicus
ORGANISM    Rattus norvegicus
REFERENCE    1 (bases 1 to 615)
AUTHORS     Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE       Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL     Genome Res. 6 (9), 791-806 (1996)
MEDLINE     9704477
PUBMED      8889548
COMMENT     Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            normalized rat brain pool library cDNA library Preparation: M.B.
            Soares Lab Clone distribution: clones will be available through
            Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=yes.

FEATURES             Location/Qualifiers
     source           1..615
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-CT0-btx-d-05-0-UI"
                     /dev_stage="ADULT"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="UI-R-CT0"
                     /note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CT0
library is a normalized library constructed from the
following rat brain tissues: embryonic day 17, embryonic
day 19, embryonic day 21, adult day 1, adult day 12, adult
day 75, adult day 200. For a detailed description of the
library from which this clone was derived, please visit
our web site at ratest.eng.uiowa.edu. The subtraction has
been previously described in (Bonaldo, Lennon and Soares,
Genome Research 6:791-806, 1996)
TAG_LIB=UI-R-CT0
TAG_TISSUE=rat brain pool
TAG_SEQ=ACTTC"
BASE COUNT        200 a 117 c 139 g 158 t 1 others
ORIGIN
Query Match      48.2%; Score 609.8; DB 10; Length 615;
Best Local Similarity 99.5%; Pred. No. 9.8e-71;
Matches 611; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      33 TTTTTCCTTTTTCCTTTCAGGATTTTAAAGTCATCATGCAAAACATACTGC 92
      |||||||
DB      2 TTTTTCCTTTTTCCTTTCAGGATTTTAAAGTCATCATGCAAAACATACTGC 61

```



```

/db xref="taxon:10090"
/clone="IMAGE:6511045"
/tissue_type="undifferentiated limb"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_134"
/notes="Vector: pCMV-SPORT6.1.cdb; Site 1: EcoRV; Site 2:
NotI; Cloned unidirectionally. Primer: Oligo dT. Average
insert size 1.7 kb. Constructed by ResGen, Invitrogen
Corp. Note: this is a NIH_MGC Library."
BASE COUNT      249 a   224 c   201 g   301 t      3 others
ORIGIN
Query Match      47.3%; Score 599.2; DB 13; Length 978;
Best Local Similarity 85.9%; Pred. No. 1.9e-69;
Matches 767; Conservative 0; Mismatches 94; Indels 32; Gaps 8;

QY 402 TTCTCACCACGCTTACTAGAAAACGAGGAGCACTAGTCAG--CTGAGTAAAGGAAG 459
Db 953 TTGCACCACCTAGTTGTTAGAAANCTAAGGAAGCCACTAAGTCTGAGTAAAGTAGG 894
QY 460 GTGAACAGGAGGAGCCTTCTACT--ATCTACAAAAAATCTCCGA-----TGCAT 511
Db 893 GCGAAACAGGAGGAGCCTTCTACTTGAGCTACCAAAAAAACAATAATCTNCGCT 834
QY 512 TATCAGAAAGATCTTATAGTACAGGTGAGATATTCCTGTTAAGAGGGGCTCTAAA 571
Db 833 TATCAGAAAGATCTTATAGTACAGGTGAGATATTCCTGTTAAGAGGGG--TCTAAA 776
QY 572 GAAAGACCTTGTAAGTTAGCAACTGTGAGGATGCCAGTTTAAATATGGAATCAACGC 631
Db 775 GAAAGACCTTGTAAGTTAGCAACTGTGAGGATGCCAGTTTAAATATGGAATCAACGC 716
QY 632 CCATCTGGGGAGGAGACACGAGGAGGAGGGGGCTCAAG-----AGAGACAC 679
Db 715 CCATCTGGGGAGGAGGAGGAGGAGGAGGGGGCTCAAG-----AGAGACAC 656
QY 680 TGATAGATCGGCCATTTGTCATCTACTGTTTACAGAAATTAACGTTAAGAGCTTTA 739
Db 655 TGATAGATCGGCCATTTGTCATCTACTGTTTACAGAAATTAACGTTAAGAGCTTTA 596
QY 740 CCGGTGACACTTTTATTCAGTTGAATTAATCTCAATGTACATGTAGTGTAAATTAATCTCT 799
Db 595 CCGGTGACACTTTTATTCAGTTGAATTAATCTCAATGTACATGTAGTGTAAATTAATCTCT 536
QY 800 ACTTCATATAGTCAAAATCTGTCTGCTCTTTGATGAGCGTGTGTTTACACTCC 859
Db 535 ACTTCATATAGTCAAAATCTGTCTGCTCTTTGATGAGCGTGTGTTTACACTCC 478
QY 860 ACCGAGCACACCCAGCTAGGAGCAAGATATCTGTTAGAGGACACAGAGGAGCAGAG 919
Db 477 ACCGAGCACACCCAGCTAGGAGCAAGATATCTGTTAGAGGACACAGAGGAGCAGAG 418
QY 920 TTCTGTTCAAAGCTGCAAGCGGTGAGTGGTATTTTAGAGAACTCACTATGAAATC 979
Db 417 TTCTGTTCAAAGCTGCAAGCGGTGAGTGGTATTTTAGAGAACTCACTATGAAATC 358
QY 980 AAGAGGACAGCTGTTTACACCATC---GTGAGTACAGTACAAAGTTACGTAATGAGCA 1036
Db 357 AAGAGGCTGAGCTGTACACTCATCACTGTGACGTACAGTACAAAGTTACGTAATGAGCA 298
QY 1037 TGGGCTGATAGTTACAGTGGTGTACATGGCAGGTGTCATTAGGAGGCTGTGCTGG 1096
Db 297 TGGGCTGATAGTTACAGTGGTGTACATGGCAGGTGTCATTAGGAGGCTGTGCTGG 238
QY 1097 TCACAGGTCTGGGAGCTACGGAGGGTCTGCACCCCTGAGCCCAAGCTCAGTCTTC 1156
Db 237 TCACAGGTCTGGGAGCTCTGGAAGGGTCTGCACCCCTGAGCTCCAGAGCTCGGTCTTC 178
QY 1157 TTAAGGACAAA--GTCCTCAACAGCTTAGTGTCTTACGTGTTCTCAGCAACGCAACT 1213
Db 177 TTAGCAACAGAGAGTCTCTCAAGAGAGTTTAGTGTCTTAGTGTCTCAGCAACGCAACT 118
QY 1214 TAGTTTCACAAGGATTTTGGCAATTTCTTAATCTGAGCAAGATATGGGATTTT 1266

```

Db 117 TAGTTCAGAAGGATTTTGGCAATTTTAAATCTGAGCAAGAAATGGGGATTTT 65

RESULT 15

CA318577/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CA318577 826 bp mRNA linear EST 26-NOV-2002
 UI-M-FWO-cbr-o-14-0-UI.r1 NIH_BMAP_FWO Mus musculus cDNA clone
 IMAGE: 6813639 5', mRNA sequence.

CA318577 GI:24536701

EST.

Mus musculus (house mouse)

Mus musculus

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 826)

NIH-MGC http://mgc.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgaabs@mail.nih.gov

Tissue procurement: Dr. Jim Lin, University of Iowa

CDNA library preparation: Dr. M. Bento Soares, University of Iowa

DNA sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

This clone was contributed by the Brain Molecular Anatomy Project

(BMAP)

Seq primer: pYX-5.

Location/Qualifiers

1. .826

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6"

/db_xref="taxon:10090"

/clone="IMAGE: 6813639"

/tissue_type="whole brain"

/dev_stage="embryo 13.5,14.5,16.5,17.5dpc"

/lab_host="DH10B (T1 phage resistant)"

/clone_lib="NIH_BMAP_FWO"

/note="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is ACCGAGACAG. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

BASE COUNT 226 a 195 c 162 g 242 t 1 others

ORIGIN

Query Match

Best Local Similarity 44.8%; Score 567; DB 14; Length 826;

Matches 682; Conservative 0; Mismatches 56; Indels 22; Gaps 6;

QY 525 TTATAGTACAGTTCAGACATATTCCTGTTAAGAGGGGCTCTTAAAGAAAGCACTTGC 584

Db 826 TTATAGTACAGTTCAGACATATTCCTGTTAAGAGGGGCTCTTAAAGAAAGCACTTGC 768

QY 585 TAAGTTAGCAACTGTGAGGATGGCCAGTTTAAATATGCACTCAACGCCCATCTCGGAG 644

Db 767 TAAGTTAG-AACTGTGAGGGGTGGCCAGTNTAAATATGCACTCAACGCCCATCTCGGAG 709

us-09-717-321a-2.rst

Qy	645	GGACACGACGGGGAAGGGGGCTCAAG-----AGAGACACTGATTAAGATCGGC	692
Db	708	GGACGCGAGTGGAGGGTGGGGGGCGGGAAGAGATGCTCAAGAGACACTGATTAAGATCGGC	649
Qy	693	CATTTCGACTACTCTGTTTTCACAGAAATTAACCGTTTAAAGAGCTTTACCGGTGACACTTT	752
Db	648	CATTGTGCCTACTGTTTTCACAGAAATTAACCGTTTAAAGAGCTTTACCGGTGACACTTT	589
Qy	753	TATTTCAGTTGAATTACTCCATGTACAAATGTAGTGTAAATTAATCTCTACTTTCATATTAGT	812
Db	588	TATTTCAGTTGAATTACTCCATGTACAAATGTAGTGTAAAGTAAATCTCTACTTTCATATTAGT	529
Qy	813	CAAAAATACTGTCTCTCTTCATGACGTCGFTGTTTCACACACTCCACCGACACACCC	872
Db	528	CAAAAATACTGTCTCTCTTCATGATGTCGFTGTTT--CACACTCCACCGACACACCC	471
Qy	873	ACCACTAGGAACAGAAATACCTTCGTTAGAGCAACACAGAGGCCAGAGTTCCTGTTCAAAGC	932
Db	470	ACAACTAGGAACAGAAATACCTTCATTAGAGGCAACACAGGAACCCAGAGTTCCTGTTCAAAGT	411
Qy	933	CTGCAGAACCGGTCTAGCTGGTATTTTAGAGAACTCAGCTATGAAATCAAGACGACAGCT	992
Db	410	CTGCAAAAGCTAGTCGCGTCTGCTTTTAGAAAACTCACTATGAAATCAAGAGCTGAGCT	351
Qy	993	GTWTACACCCATC--GTGACGTACAGTACAAAGTTAGTGAATCAGAGATGGCTGATAAGT	1049
Db	350	GTCACTCTCATCTACTGTGACGTACAGTACAAAGTTTACGTAATGAACATGGGCTGATAGT	291
Qy	1050	TACAGGTGCGCTTACATGGCACGCTGTCAITTAAGGAGGCTGTGCTGTGTCAACGGTCTGG	1109
Db	290	TACAGGTGCGCTTACATGGCAACGTGTCAITTAAGGAGGCTGTGCTGTGTCAACGGTCTGG	231
Qy	1110	GAGCTACGGAGGGTCTGCAACCCCTGAGCCAGAGCTGCAGTCTTCTTAAGGACAAA--	1167
Db	230	GAACTCTGGAAGGGTCTGCAACCCCTGGCTCCAGAGCTGGGTCTTCTTAGCAACAGAAG	171
Qy	1168	-GTCTCTCAACAGCTTAGTGTCTTACGTGTTCTCAGCAACACGCAACTTAGTTCACAAGGT	1226
Db	170	TCCTCAAGAGAGTTTAGTGTCTTAGTGTCTCAGCAACACGCAACTTAGTTCAGAAGGT	111
Qy	1227	ATTTTGGCAATCTTAATCTGAGCAAGAAATAGGGGATTTT	1266
Db	110	ATTTTGGCAATTTTAAATCTGAGCAAGAAATGGGGATTTT	71

Search completed: November 23, 2003, 15:50:01
Job time : 2984.07 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 23, 2003, 10:43:52 ; Search time 818.065 Seconds
(without alignments)
10338.975 Million cell updates/sec

Title: US-09-717-321A-17

Perfect score: 348

Sequence: 1 tgaacatcactgtctgtgc.....ctttggaacttgaaaaaaa 348

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gsl:*
29: gb_gss2:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	324	93.1	429	9	AI059212 UI-R-CI-1
C 2	323.4	92.9	493	10	BF554834 UI-R-E0-C
C 3	323.4	92.9	687	12	BM389059 UI-R-D20-
C 4	321.8	92.5	459	13	BQ780699 UI-R-PF0-

C 5	321.8	92.5	640	13	BQ780117
C 6	321.4	92.4	477	9	AW520555
C 7	321.4	92.4	489	9	AA859879
C 8	321.4	92.4	565	12	BI273986
C 9	321.4	92.4	760	13	BQ191985
C 10	320.8	92.2	472	13	BQ190077
C 11	320.8	92.2	615	10	BQ380448
C 12	320.8	92.2	639	10	BE111691
C 13	320.8	92.2	705	13	BQ200122
C 14	318.2	91.4	465	9	AA899757
C 15	317	91.1	405	10	BE104696
C 16	317	91.1	412	13	BQ780657
C 17	317	91.1	517	9	AA859940
C 18	317	91.1	636	13	BQ780658
C 19	317	91.1	638	12	BI283790
C 20	313.8	90.2	469	12	BI276611
C 21	312.4	89.8	557	10	BG665005
C 22	310.6	89.3	446	12	BI286743
C 23	305.4	87.8	463	10	BF548903
C 24	304.8	87.6	316	9	AI412434
C 25	300	86.2	316	9	AI409795
C 26	297.2	85.4	407	9	AI009011
C 27	297.2	85.4	455	9	AI012196
C 28	295.6	84.9	553	9	AA799542
C 29	293.6	84.4	427	9	AI178082
C 30	292	83.9	431	9	AI410870
C 31	285.6	82.1	292	9	AI598992
C 32	278	79.9	294	9	AI236740
C 33	276.8	79.5	522	14	C06842
C 34	273.8	78.7	277	10	BF420446
C 35	251.2	72.2	277	10	BF420436
C 36	246.4	70.8	404	10	BF549054
C 37	243.4	69.9	644	14	CA339407
C 38	238.4	68.5	256	9	AI234716
C 39	227.6	65.4	394	12	BM022196
C 40	226.6	65.1	322	13	BQ757481
C 41	226.6	65.1	324	13	BQ756887
C 42	226.6	65.1	546	12	BQ033391
C 43	226.6	65.1	655	10	B2996534
C 44	226	64.9	408	12	BI157798
C 45	225.8	64.9	404	12	BM502810

ALIGNMENTS

RESULT 1
AI059212/c
LOCUS
DEFINITION
UI-R-CI-lb-e-11-0-UI-sl UI-R-CI Rattus norvegicus cDNA clone
UI-R-CI-lb-e-11-0-UI 3', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 429)
AUTHORS
TITLE
Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL
MEDLINE
PUBMED
COMMENT
Genome Res. 6 (9), 791-806 (1996)
8889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu

The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized adult lung library. cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics

Seq primer: M13 Forward.

Location/Qualifiers

1. 429
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-C1-1b-e-11-0-UI"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-C1"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-C1 library is a subtracted library derived from the UI-R-C0 library, which is a subtracted library derived from the UI-R-A1 and UI-R-E1 libraries. The UI-R-A1 library consisted of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, and muscle. The UI-R-E1 library consisted of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-C1) was constructed as follows: PCR amplified cDNA inserts from UI-R-C0 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the UI-R-C0 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-C1 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)."

BASE COUNT 130 a 87 c 91 g 121 t

Query Match 93.1%; Score 324; DB 9; Length 429;
Best Local Similarity 95.7%; Pred. No. 4.5e-37;
Matches 333; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 TGAACATCAGTGTCTTCCAGAGTACCAACACTGTCATGTGATTCGCGCCCTCT 60
Db 357 TGAACATCAGTGTCTTCCAGAGTACCAACACTGTCATGTGATTCGCGCCCTAT 298

QY 61 AGACCTCACCAGCGGACACATGCTCCGGTACCTTTCGGTCTGTGAGTTCTGCAAG 120
Db 297 AGACCTCACCAGCGGACACATGCTCCGGTACCTTTCGGTCTGTGAGTTCTGCAAG 238

QY 121 CGCTAGTGTCTAACCGCGTCTGTGTACAACTTAACCTGCGCAAGAACACAGTGTGGGCC 180
Db 237 CGCTAGTGTCTAACCGCGTCTGTGTACAACTTAACCTGCGCAAGAACACAGTGTGGGCC 178

QY 181 TTTCGACCACTAGAACAACTTTTTCATATGACAGTTCGAGATTCGAGTGTGTTTA 240
Db 177 TTTCGACCACTAGAACAACTTTTTCATATGACAGTTCGAGATTCGAGTGTGTTTA 118

QY 241 CATTTGATCTTTTCTAATGAGTTCAGTATCTTTTTCATATGATGATGATTAATAATCT 300
Db 117 CATTTGATCTTTTCTAATGAGTTCAGTATCTTTTTCATATGATGATGATTAATAATCT 58

QY 301 TGAATCATATAAAAAAATAAATGCTTTTGGAACTTGAATAAAAA 348
Db 57 TGAATCATATAAACTGGTTCATGCTTGTGCTTTTGGAACTTGAATAAAAA 10

RESULT 2

BF554834

LOCUS

DEFINITION

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

1 (Bases 1 to 493)

Bonaldo, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

8889548

Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: bento-soares@uiowa.edu

cDNA Library Preparation: M.B. Soares Lab Clone distribution:

clones will be available through Research Genetics (www.resgen.com)

This clone is also available through the I.M.A.G.E. Consortium at

LLNL (info@image.llnl.gov). IMAGE ID= 1770064 The following

repetitive elements were found in this cDNA sequence: 463-485,

>AT rich#Low complexity

Seq primer: M13 Forward.

Location/Qualifiers

1. 493

/organism="Rattus norvegicus"

/mol_type="mRNA"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="UI-R-E0-cc-c-06-0-UI"

/dev_stage="embryonic"

/lab_host="DH10B (Life Technologies)"

/clone_lib="UI-R-E0"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site 1: NotI; Site 2: EcoRI; This library

consists of a mixture of individually tagged normalized

libraries constructed from 8, 12 and 18-day embryo. The

tag is a string of 3-5 nucleotides present between the

Not I site and the oligo-dT track which allows

identification of the library of origin of a clone within

the mixture."

BASE COUNT 130 a 109 c 95 g 159 t

ORIGIN

Query Match 92.9%; Score 323.4; DB 10; Length 493;

Best Local Similarity 99.7%; Pred. No. 5.1e-37;

Matches 324; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGAACATCAGTGTCTTCCAGAGTACCAACACTGTCATGTGATTCGCGCCCTCT 60

Db 157 TGAACATCAGTGTCTTCCAGAGTACCAACACTGTCATGTGATTCGCGCCCTCT 216

QY 61 AGACCTCACCAGCGGACACATGCTCCGGTACCTTTCGGTCTGTGAGTTCTGTCAG 120

Db 217 AGACCTCACCAGCGGACACATGCTCCGGTACCTTTCGGTCTGTGAGTTCTGTCAG 276

QY 121 CGCTAGTGTCTAACCGCGTCTGTGTACAACTTAACCTGCGCAAGAACACAGTGTGGGCC 180

Db 277 CGCTAGTGTCTAACCGCGTCTGTGTACAACTTAACCTGCGCAAGAACACAGTGTGGGCC 336

QY 181 TTTCGACCACTAGAACAACTTTTTCATATGACAGTTCGAGATTCGAGTGTGTTTA 240

Db
 337 TTTGACCACTAGAACAACTTTTCAATTGACAGTTGCGAATTGTGGAGTGTTTTA 396
 QY
 241 CATTTGATCTTTTGTAAATGCACTTAGCAGTATGTTTTGCATGTATGACTTAAATAATCCT 300
 Db
 397 CATTTGATCTTTTGTAAATGCACTTAGCAGTATGTTTTGCATGTATGACTTAAATAATCCT 456
 QY
 301 TGAATCATATAAAAAAAAAAAAAA 325
 Db
 457 TGAATCATATAAAAAAAAAAAAAA 481

RESULT 3	BM389059/c	EM389059	687 bp	linear	EST 17-JAN-2002
LOCUS		UI-R-DZ0-cko-j-07-0-UI.s1		UI-R-DZ0	Rattus norvegicus cDNA clone
DEFINITION		UI-R-DZ0-cko-j-07-0-UI 3', mRNA sequence.			

ACCESSION	BM389059
VERSION	BM389059.1
KEYWORDS	GI:18189112
SOURCE	EST.
ORGANISM	<i>Rattus norvegicus</i> (Norway rat)
	<i>Rattus norvegicus</i>
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; <i>Rattus</i> .

REFERENCE	AUTHORS	TITLE	JOURNAL
1	(Bases 1 to 687)	Bonaldo, M.F., Lennon, G. and Soares, M.B.	
		Normalization and subtraction: two approaches to facilitate gene discovery	Genome Res. 6 (9), 791-806 (1996)

97044477
MEDLINE
889548
PUBLISHED
COMMENT
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@iowa.edu

email: denno-soares@uowa.edu.au

The sequence contained an oligo-dr track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dr track served to verify it as a clone from the non-normalized cartilaginous tumor library cDNA library

Preparation: M.B. Soares Lab Clone Distribution: clones will be available through Research Genetics (www.resgen.com)

Seq primer: M13 Forward

POLYA=Yes.

```

FEATURES
source
Location/Qualifiers
1..687
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-D20-cko-j-07-0-UI"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-D20"
/note="Vector: pVT3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; UI-R-D20 is a
non-normalized Rat cartilaginous tumor library (RCT)
constructed in pBstI7 PAC vector according to the procedure
described by Honaldo, Lennon & Soares (Genome Research
Genome 6: 791-806, 1996). The oligonucleotide used to
prime first strand synthesis contained the sequence tag
CATCTCTGTA between the Not I cloning site and d118
stretch. The Rat cartilaginous tumor tissue was provided
by Dr Jeff Stevens at the University of Iowa.
TAG_LIB=UI-R-D20
TAG_TISSUE=cartilaginous tumor
TAG_SEQ=CATCTCTGTA"
222 a 130 c 163 g 172 t

```

ORIGIN		Query Match Best Local Similarity Matches 324; Conservative 0; Mismatches 1; Indels 0; Gaps 0
QY	1	TGAACATCACTGCTTTGCCAGAGTACCAACACGTGCATGTGATTGATCGCCGCCCCCCT 60
Db	326	TGAACAATCACTGCTTTGCCAGAGTACCAACACGTGCATGTGATTGATCGCCGCCCCCCT 267
QY	61	AGACTCACCAACGGGACACATCCTTCGGTGACTTTGGGTCTGTGAGGTTCTGTCAAG 120
Db	266	AGACTCACCAACGGGACACATCCTTCGGTGACTTTGGGTCTGTGAGGTTCTGTCAAG 207
QY	121	CGTAGTGTCTAAGCCGCTTCCTGTACAACTTAACCTGCGCAAGAACACACAGTGTGGGCC 180
Db	206	CGTAGTGTCTAAGCCGCTTCCTGTACAACTTAACCTGCGCAAGAACACACAGTGTGGGCC 147
QY	181	TTTCGACCCTAGAACAACCTTTTTCAATTGCAGTTGCAGAAATGTGGAGTGTTTTTA 240
Db	146	TTTCGACCCTAGAACAACCTTTTTCAATTGCAGTTGCAGAAATGTGGAGTGTTTTTA 87
QY	241	CATTGATCTTTGCTAATGCAGTTAGCAGTATGTTTTGCATGTATGACTTAATAAATCCT 300
Db	86	CATTGATCTTTGCTAATGCAGTTAGCAGTATGTTTTGCATGTATGACTTAATAAATCCT 27
QY	301	TGAATCATAAAAAIAAAAAAAAAA 325
Db	26	TGAATCATAAAAAIAAAAAAAAAA 2

RESULT 4	
BQ780699/c	
LOCUS	BQ780699 459 bp mRNA linear EST 26-JUL-2002
DEFINITION	UI-R-FFO-cpc-k-22-0-UI.s1 UI-R-FFO Rattus norvegicus cDNA clone
	UI-R-FFO-cpc-k-22-0-UI 3' , mRNA sequence.

ACCESSION BQ780699
 VERSION BQ780699.1 GI:21989171
 KEYWORDS EST.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 REFERENCE 1 (bases 1 to 459)
 AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE
 PUBMED
 COMMENT

97044477
 8889548
 Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road , 4156 MERRF, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: bento-soares@uiowa.edu
 Tissue Procurement: Jeff Stevens
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: DISTRIBUTION: Researchers may obtain clones
 from Research Genetics (www.resgen.com).
 Seq primer: M13 FORWARD

```

FEATURES
    source
    Location/Qualifiers
        1. .459
            /organism="Rattus norvegicus"
            /mol_type="mRNA"
            /db_xref="taxon:10116"
            /clone="UI-R-EF0-cpc-k-22-0-UI"
            /tissue type="Mixed tissues"

```

```

/dev stage="Adult"
/lab host="DH10B (Life Technologies) (T1 phage resistant)"
/clone lib="UI-R-PF0"
/note="Vector: pT73-Pac (Pharmacia) with a modified
polylinker; Site 1: Ecor I; Site 2: Not I; UI-R-PF0 is a
subtracted cDNA library containing the following tissue(s)
): Normal cartilage and SR-JWS Tumor line. The
subtraction was made according to Bonaldo, Lennon and
Soares, Genome Research, 6:791-806, 1996. The
oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for these libraries are: CTAATGGAGC,
CATCTTCTGA.
TAG LIB=UI-R-PF0
TAG_TISSUE=cartilage
TAG_SEQ=CTAATGGAGC"
BASE COUNT 145 a 91 c 103 g 120 t
ORIGIN

```

```

Query Match 92.5%; Score 321.8; DB 13; Length 459;
Best Local Similarity 99.4%; Pred. No. 9e-37;
Matches 323; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAACATCATCTGTCTTCCAGAGTACCAACACTGTCATGTCATGTCGCGCCCTCT 60
Db 325 TGAACATCATCTGTCTTCCAGAGTACCAACACTGTCATGTCATGTCGCGCCCTCT 266

QY 61 AGACCTCACCCACGGGACACATGCTTCGGTACCTTTGGGTCCTGAGGTCCTCAAG 120
Db 265 AGACCTCACCCACGGGACACATGCTTCGGTACCTTTGGGTCCTGAGGTCCTCAAG 206

QY 121 CGCTAGTGTCTAACGCGGTTCTGTACAACTCACTCACTGCGAAGACACAGTGTGGGCC 180
Db 205 CGCTAGTGTCTAACGCGGTTCTGTACAACTCACTCACTGCGAAGACACAGTGTGGGCC 146

QY 181 TTTCGACCACTAGACAAACTTTTTCATTTGACAGTTCGCAATTTGGAGTGTCTTTA 240
Db 145 TTTCGACCACTAGACAAACTTTTTCATTTGACAGTTCGCAATTTGGAGTGTCTTTA 86

QY 241 CATTGATCTTTGCTAATGCTAGTACGATGTTTTCATGATGATGATGATGATGATGAT 300
Db 85 CATTGATCTTTGCTAATGCTAGTACGATGTTTTCATGATGATGATGATGATGATGAT 26

QY 301 TGAATCATATAAAAAAAAAAAAAA 325
Db 25 TGAATCATATAAAAAAAAAAAAAA 1

```

```

RESULT 5
BQ780117/c
LOCUS
DEFINITION
UI-R-PF0-cow-j-10-0-UI.sl mRNA linear EST 26-JUL-2002
UI-R-PF0-cow-j-10-0-UI.sl UI-R-PF0 Rattus norvegicus cDNA clone
BQ780117
VERSION
BQ780117.1 GI:21988589
KEYWORDS
EST.
SOURCE
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 640)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
AUTHORS
Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL
Genome Res. 6 (9), 791-806 (1996)
MEDLINE
97044477
PUBMED
889548
COMMENT
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA

```

Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Jeff Stevens
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: DISTRIBUTION: Researchers may obtain clones
from Research Genetics (www.resgen.com).
Seq primer: ML3 FORWARD
POLYA-Yes.

FEATURES
source
Location/Qualifiers
1..640
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="UI-R-PF0-cow-j-10-0-UI"
/tissue_type="Mixed tissues"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-R-PF0"
/note="Vector: pT73-Pac (Pharmacia) with a modified
polylinker; Site 1: Ecor I; Site 2: Not I; UI-R-PF0 is a
subtracted cDNA library containing the following tissue(s)
): Normal cartilage and SR-JWS Tumor line. The
subtraction was made according to Bonaldo, Lennon and
Soares, Genome Research, 6:791-806, 1996. The
oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for these libraries are: CTAATGGAGC,
CATCTTCTGA.

TAG LIB=UI-R-PF0
TAG_TISSUE=rat SRC-JWS tumor line
TAG_SEQ=CATCTTCTGA"
BASE COUNT 208 a 121 c 152 g 159 t
ORIGIN

Query Match 92.5%; Score 321.8; DB 13; Length 640;
Best Local Similarity 99.4%; Pred. NO. 7.5e-37;
Matches 323; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 TGAACATCATCTGTCTTCCAGAGTACCAACACTGTCATGTCATGTCGCGCCCTCT 60
Db 325 TGAACATCATCTGTCTTCCAGAGTACCAACACTGTCATGTCATGTCGCGCCCTCT 267

QY 61 AGACCTCACCCACGGGACACATGCTTCGGTACCTTTGGGTCCTGAGGTCCTCAAG 120
Db 266 AGACCTCACCCACGGGACACATGCTTCGGTACCTTTGGGTCCTGAGGTCCTCAAG 207

QY 121 CGCTAGTGTCTAACGCGGTTCTGTACAACTCACTCACTGCGAAGACACAGTGTGGGCC 180
Db 206 CGCTAGTGTCTAACGCGGTTCTGTACAACTCACTCACTGCGAAGACACAGTGTGGGCC 147

QY 181 TTTCGACCACTAGACAAACTTTTTCATTTGACAGTTCGCAATTTGGAGTGTCTTTA 240
Db 146 TTTCGACCACTAGACAAACTTTTTCATTTGACAGTTCGCAATTTGGAGTGTCTTTA 87

QY 241 CATTGATCTTTGCTAATGCTAGTACGATGTTTTCATGATGATGATGATGATGATGAT 300
Db 86 CATTGATCTTTGCTAATGCTAGTACGATGTTTTCATGATGATGATGATGATGATGAT 27

QY 301 TGAATCATATAAAAAAAAAAAAAA 325
Db 26 TGAATCATATAAAAAAAAAAAAAA 2

```

```

RESULT 6
AWS20555/c
LOCUS
DEFINITION
UI-R-BJOp-afx-g-09-0-UI.sl UI-R-BJOp Rattus norvegicus cDNA clone
UI-R-BJOp-afx-g-09-0-UI.3', mRNA sequence.
ACCESSION
AWS20555

```


Tel: 319 335 8250
Fax: 319 335 9565

Email: bento-soares@uiowa.edu
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized osteoblast library cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES

Source 1. .760
Location/Qualifiers
/organism="Rattus norvegicus"
/mol_type="mrna"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-DRI-cky-h-11-0-UI"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-DRI"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-DRI library is a normalized Rat Osteoblast library (nREO) constructed in pT377 vector according to the procedure described by Bonaldo, Lennon & Soares (Normalization and Subtraction: Two Approaches to Facilitate Gene Discovery. Genome Research 6: 791-806, 1996). The oligonucleotide used to prime first strand synthesis contained the sequence tag AAGATATCAA between the Not I cloning site and dT18 stretch. The Rat Osteoblast tissue was provided by Lian & Stein of the University of Massachusetts Medical School.
TAG LIB=UI-R-DRI
TAG TISSUE=osteoblast
TAG SEQ=AAGATATCAA"
BASE COUNT 244 a 173 g 198 t 2 others
ORIGIN

Query Match 92.4%; Score 321.4; DB 13; Length 760;
Best Local Similarity 99.4%; Pred. No. 7.7e-37;
Matches 322; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAACATCACTCTCTGCGAGATACCACTGTCATGTGATGCGGCCCTCT 60
Db |||||||
324 TGAACATCACTCTCTGCGAGATACCACTGTCATGTGATGCGGCCCTCT 265
QY 61 AGACCTCACCCGCGGACACATGCTTCGGTACCTTTGGTGTGTCAGGTTCTGTCAAG 120
Db |||||||
264 AGACCTCACCCGCGGACACATGCTTCGGTACCTTTGGTGTGTCAGGTTCTGTCAAG 205
QY 121 CGCTAGTGTACCGCGTCTGTACAACTTACTCACTGCGAAGACACAGTGTGGGCC 180
Db |||||||
204 CGCTAGTGTACCGCGTCTGTACAACTTACTCACTGCGAAGACACAGTGTGGGCC 145
QY 181 TTTCGACCACTAGACAAACTTTTTCATTTGACAGTTCGAGATTTGTGAGTGTTTTA 240
Db |||||||
144 TTTCGACCACTAGACAAACTTTTTCATTTGACAGTTCGAGATTTGTGAGTGTTTTA 85
QY 241 CATTTGATCTTTTCTAATGAGTTAGCAGTATCTTTTTCATGATGATCACTTAATAATCT 300
Db |||||||
84 CATTTGATCTTTTCTAATGAGTTAGCAGTATCTTTTTCATGATGATCACTTAATAATCT 25
QY 301 TGAATCATAAAAAATAAAAAA 324
Db |||||||
24 TGAATCATAAAAAATAAAAAA 1

RESULT 10
BQ190077/c
LOCUS 472 bp mrna linear EST 30-APR-2002

DEFINITION

UI-R-CNI-cjs-h-07-0-UI-s3 UI-R-CNI Rattus norvegicus cDNA clone
UI-R-CNI-cjs-h-07-0-UI 3', mRNA sequence.

ACCESSION

BQ190077

VERSION

BQ190077.1 GI:20365628

KEYWORDS

EST.

SOURCE

Rattus norvegicus (Norway rat)

ORGANISM

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE

1 (bases 1 to 472)

AUTHORS

Bonaldo, M.F., Lennon, G. and Soares, M.B.

TITLE

Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL

Genome Res. 6 (9), 791-806 (1996)

MEDLINE

97044477

PUBMED

889548

COMMENT

Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized rat eye library cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES

Location/Qualifiers

1. .472

/organism="Rattus norvegicus"

/mol_type="mrna"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="UI-R-CNI-cjs-h-07-0-UI"

/dev_stage="adult"

/lab_host="DH10B (Life Technologies)"

/clone_lib="UI-R-CNI"

/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-CNI library is a subtracted library derived from the following pool of seven normalized rat libraries: normalized rat seminal vesicles, normalized rat penis, normalized rat bladder, normalized rat cervix, normalized rat adipose, normalized rat fundus, and normalized rat salivary gland. It was constructed according to the procedure described by Bonaldo, Lennon & Soares (Genome Research 6: 791-806, 1996). For construction of the CNI library, plasmid DNA from the pool of normalized libraries was electroporated into competent bacteria for the production of single-stranded circular DNA. This was then used as a tracer in a subtractive hybridization with a driver (PCR amplified inserts from a plasmid DNA template preparation) comprising: a) a pool of about 34,000 clones from the Rat Unigene Set corresponding to plates R-5-AA-NN excluding plates R-5-MM and MN. This pool represented 40% of the final driver population. b) a pool of about 29,000 clones from subtracted libraries CA0 and CA1 corresponding to plates R-CA0-ANV through R-CA0-AXS, R-CA0-AZX through R-CA0-BAZ, R-CA0-BFE through R-CA0-BHY, R-CA0-BJS, R-CA0-BKE, R-CA0-BKG-H, R-CA0-BKJ-K, R-CA0-BKP through R-CA0-BKS, R-CA0-BKU-V, R-CA0-BLY through R-CA0-BMA, R-CA0-BMC through R-CA0-BME, R-CA0-BNS, R-CA0-BOB through R-CA0-BOJ, R-CA0-BPA through R-CA0-BPG, R-CA1-BBA through R-CA1-BDA, R-CA1-BHJ through R-CA1-BJF, R-CA1-BKJ, R-CA1-BKT through R-CA1-BKB, R-CA1-BKD, R-CA1-BKF, R-CA1-BKI, R-CA1-BKT, R-CA1-BLF, R-CA1-BLH through

R-CA1-BLN, R-CA1-BLS, R-CA1-BLU-V, R-CA1-BNR, and R-CA1-BLE. The resulting pool represented 20% of the final driver population. c) a pool of about 15,000 clones from non-normalized libraries CS0s, CT0s, CU0s, CW0s, CX0s and normalized libraries CS0, CT0, CU0, CW0, and CX0 corresponding to plates R-CS0s-CBD through R-CS0s-CBO, R-CT0s-CAM through R-CT0s-CAX, R-CU0s-CBP through R-CU0s-CCA, R-CW0s-CCB through R-CW0s-CCM, R-CX0s-CCN through R-CX0s-CCX, R-CS0-BSD, R-CS0-BTD through R-CS0-BTV, R-CS0-BVM, R-CT0-BTW through R-CT0-BUP, R-CT0-BVN, R-CU0-BUQ through R-CU0-BVL, R-CW0-BVV through R-CW0-BWP, R-CX0-BXN through R-CW0-BXO, R-CX0-BWQ through R-CX0-BXM. The resulting pool represented 5% of the final driver population. d) a pool of about 5,000 clones (1,000 from non-normalized eye library CV0 and 4,000 from normalized eye library CV1) corresponding to plates R-CV0-BRH through R-CV0-BRR, R-CV1-BRS through R-CV1-BSC, R-CV1-BSE through R-CV1-BTC, and R-CV1-BVO through R-CV1-BVU. This pool represented about 5% of the final driver population. e) A pool of about 10,000 clones from subtracted library BS2, BV0 and BV0p (7-9.5 kb cDNA library fraction from rat whole embryo), and BX0 (0.5-7kb cDNA library fraction from rat whole embryo) corresponding to plates R-BS2-BDB through R-BS2-BFB, R-BV0-ANK through R-BV0-ANR, R-BV0p-AOI through R-BV0p-AOX, and R-BX0-AQY through R-BX0-ASH. The resulting pool represented 5% of the final driver population. f) a pool of about 7,000 clones from the seven non-normalized libraries that make up the tracer including CV0, CZ0, DAO, DB0, DCO, DD0, and DE0 corresponding to plates R-CY0-BXP through R-CY0-BXZ, R-CZ0-BYA through R-CZ0-BYI, R-CZ0-BZB-C, R-DAO-BYJ through R-DAO-BYP, R-DAO-BZD through R-DAO-BZH, R-DB0-BYQ through R-DB0-BZA, R-DC0-BZI through R-DC0-BZQ, R-DC0-CAY through R-DC0-CBA, R-DD0-BZR through R-DD0-CAA, R-DD0-CBB-C, and R-DE0-CAB through R-DE0-CAL. The resulting pool represented about 10% of the final driver population. g) a pool of about 2,000 clones from the pool of normalized libraries, CN0, that makes up the tracer. The corresponding plates are R-CN0-BKW through R-CN0-BLD, R-CN0-BLG, R-CN0-BLP through R-CN0-BLR, R-CN0-BLT, R-CN0-BLM-X, R-CN0-BMS, and R-CN0-BMF through R-CN0-BML. This pool represented 5% of the final driver population. h) a pool of the 28 most abundant clones in the CN0 pool corresponding to the following addresses: bkx-a-09-0-UI, bkx-b-09-0-UI, bkx-b-11-0-UI, bkx-b-10-0-UI, bkx-d-01-0-UI, bkx-d-06-0-UI, bkx-g-08-0-UI, bkx-h-12-0-UI, bkx-a-05-0-UI, bkx-a-06-0-UI, bkx-a-11-0-UI, bkx-c-06-0-UI, bkx-c-09-0-UI, bkx-d-10-0-UI, bla-a-01-0-UI, bla-a-02-0-UI, bla-f-04-0-UI, bla-g-07-0-UI, bla-g-12-0-UI, blb-a-12-0-UI, blb-f-02-0-UI, blc-a-11-0-UI, blc-e-95-0-UI, bld-1-08-0-UI, bld-f-02-0-UI, blg-h-04-0-UI, blr-a-05-0-UI, blt-f-08-0-UI. This pool represented 5% of the final driver population. i) One abundant CN0 clone (corresponding to the address bkx-a-11-0-UI) was digested with Not I and Eco RI and the resulting insert was gel purified. This purified insert was added directly to the driver so that it represented 5% of the final driver population.

BASE COUNT 154 a 102 g 123 t
ORIGIN
TAG LIB=UI-R-CN1
TAG TISSUE=rat eye
TAG SEQ=CAGCC"

Query Match 92.2%; Score 320.8; DB 13; Length 472;
Best Local Similarity 99.4%; Pred. No. 1.2e-36;
Matches 322; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TGAACATCACTGTTGCCAGAGTACCAACACTGTCATGTCATGATGCGGCCCTCT 60
Db 324 TGAACATCACTGTTGCCAGAGTACCAACACTGTCATGTCATGATGCGGCCCTCT 265
QY 61 AGACCTCACCCACGCGGACACATGCTCCGGTACCTTTGGGTCTGTGAGGTTCTGTCAAG 120

Db 264 AGACCTCACCCACGCGGACACATGCTCCGGTACCTTTGGGTCTGTGAGGTTCTGTCAAG 205
QY 121 CGCTAGTCTCTAAGCGCGTTCTGTCAACCTCACTGCGCAAGAACACAGTGTGGGCC 180
Db 204 CGCTAGTCTCTAAGCGCGTTCTGTCAACCTCACTGCGCAAGAACACAGTGTGGGCC 145
QY 181 TTTCGACCACCTAGAACAAACTTTTTCATTTGACAGTTGACAGTTGCGAGATGTTGTTTA 240
Db 144 TTTCGACCACCTAGAACAAACTTTTTCATTTGACAGTTGACAGTTGCGAGATGTTGTTTA 85
QY 241 CATTGATCTTTTCTTAATGCGAGTTAGCAGTATGTTTTCATGTCATGACTTAATAATCT 300
Db 84 CATTGATCTTTTCTTAATGCGAGTTAGCAGTATGTTTTCATGTCATGACTTAATAATCT 25
QY 301 TGAATCATATAAAAAA 324
Db 24 TGAATCATATAAAAAA 1

RESULT 11
BG380448/c

LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BG380448 615 bp mRNA linear EST 12-MAR-2001
UI-R-CT0-btx-d-05-0-UI.s1 UI-R-CT0 Rattus norvegicus cDNA clone
UI-R-CT0-btx-d-05-0-UI 3', mRNA sequence.
BG380448
BG380448.1 GI:13304920
EST.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

1 (bases 1 to 615)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
8889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu

The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
normalized rat brain pool library cDNA Library Preparation: M.B.
Soares Lab Clone distribution: clones will be available through
Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES
source

1..615
Location/Qualifiers
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CT0-btx-d-05-0-UI"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-CT0"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-CT0
library is a normalized library constructed from the
following rat brain tissues: embryonic day 17, embryonic
day 19, embryonic day 21, adult day 1, adult day 12, adult
day 75, adult day 200. For a detailed description of the

library from which this clone was derived, please visit our web site at ratest.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)

TAG_L1B=UI-R-BJ1
TAG_TISSUE=rat brain pool
TAG_SEQ=ACATTC

BASE COUNT 200 a 117 c 139 g 158 t 1 others
ORIGIN

Query Match 92.2%; Score 320.8; DB 10; Length 615;
Best Local Similarity 99.1%; Pred. No. 1.1e-36;
Matches 322; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAACATCACTGCTTCCAGAGTACCAACACTGCTATGATGATGCGGCCCTCT 60
Db 326 TGAACATCACTGCTTCCAGAGTACCAACACTGATGATGATGCGGCCCTCT 267
QY 61 AGACTCACCCACCGGACACATGCTTCCGTACCTTTGGGTCTGTGAGGTTCTGTCAAG 120
Db 266 AGACTCACCCACCGGACACATGCTTCCGTACCTTTGGGTCTGTGAGGTTCTGTCAAG 207
QY 121 CGCTAGTGTAAACCGCTTCTGTACAACTTAACCTACCTGCGCAAGACACAGTGTGGGCC 180
Db 206 CGCTAGTGTAAACCGCTTCTGTACAACTTAACCTACCTGCGCAAGACACAGTGTGGGCC 147
QY 181 TTTCGACCCTAGAACAACTTTTTCATTAATGACAGTTCAGAAATTTGGAGTGTTTTA 240
Db 146 TTTCGACCCTAGAACAACTTTTTCATTAATGACAGTTCAGAAATTTGGAGTGTTTTA 87
QY 241 CATTGATCTTTTGTGTAATGAGTGTAGAGTATGTTTGCATGTATGACTTAAATAATCCT 300
Db 86 CATTGATCTTTTGTGTAATGAGTGTAGAGTATGTTTGCATGTATGACTTAAATAATCCT 27
QY 301 TGAATCATATAAAAAA 325
Db 26 TGAATCATATAAAAAA 2

RESULT 12
BE111691/c

LOCUS BE111691 639 bp mRNA linear EST 13-JUN-2000
DEFINITION UI-R-BJ1-avv-f-10-0-UI.s1 UI-R-BJ1 Rattus norvegicus cDNA clone
UI-R-BJ1-avv-f-10-0-UI 3', mRNA sequence.

ACCESSION BE111691
VERSION BE111691.1 GI:8503796
KEYWORDS EST
SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 639)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene discovery
Genome Res. 6 (9), 791-806 (1996)
9704477

JOURNAL MEDLINE
PUBMED 889548

COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized ventricle at 13 dpc library cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available

through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES
source

Location/Qualifiers
1..639
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BJ1-avv-f-10-0-UI"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-BJ1"

/note="vector: p773D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-BJ1 library is a subtracted library derived from the following tissues: atrium at 16.5 dpc, ventricle at 16.5 dpc, AV canal at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc, AV canal at 15 dpc, ventricle at 13 dpc, and adult heart. For a detailed description of the library from which this clone was derived, please visit our web site at ratest.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
TAG_L1B=UI-R-BJ1
TAG_TISSUE=ventricle at 13 dpc
TAG_SEQ=CAGCGA"

BASE COUNT 209 a 121 c 151 g 157 t 1 others
ORIGIN

Query Match 92.2%; Score 320.8; DB 10; Length 639;
Best Local Similarity 99.1%; Pred. No. 1e-36;
Matches 322; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAACATCACTGCTTCCAGAGTACCAACACTGCTATGATGATGCGGCCCTCT 60
Db 325 TGAACATCACTGCTTCCAGAGTACCAACACTGCTATGATGATGCGGCCCTCT 266
QY 61 AGACTCACCCACCGGACACATGCTTCCGTACCTTTGGGTCTGTGAGGTTCTGTCAAG 120
Db 265 AGACTCACCCACCGGACACATGCTTCCGTACCTTTGGGTCTGTGAGGTTCTGTCAAG 206
QY 121 CGCTAGTGTAAACCGCTTCTGTACAACTTAACCTACCTGCGCAAGACACAGTGTGGGCC 180
Db 205 CGCTAGTGTAAACCGCTTCTGTACAACTTAACCTACCTGCGCAAGACACAGTGTGGGCC 146
QY 181 TTTCGACCCTAGAACAACTTTTTCATTAATGACAGTTCAGAAATTTGGAGTGTTTTA 240
Db 145 TTTCGACCCTAGAACAACTTTTTCATTAATGACAGTTCAGAAATTTGGAGTGTTTTA 86
QY 241 CATTGATCTTTTGTGTAATGAGTGTAGAGTATGTTTGCATGTATGACTTAAATAATCCT 300
Db 85 CATTGATCTTTTGTGTAATGAGTGTAGAGTATGTTTGCATGTATGACTTAAATAATCCT 26
QY 301 TGAATCATATAAAAAA 325
Db 25 TGAATCATATAAAAAA 1

RESULT 13
BQ200122/c

LOCUS BQ200122

DEFINITION UI-R-EB1-clf-i-07-0-UI.s1 UI-R-EB1 Rattus norvegicus cDNA clone
UI-R-EB1-clf-i-07-0-UI 3', mRNA sequence.

ACCESSION BQ200122

VERSION BQ200122.1 GI:20416587

KEYWORDS EST.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 705)
Bonaldo,M.F., Lennon,G. and Soares,M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 PUBMED 8889548
 COMMENT Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: bento-soares@uiowa.edu
 The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized duodenum library cDNA Library Preparation: M.B. Soares Lab Clone distribution: Clones will be available through Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLYA=Yes.

FEATURES

source
 Location/Qualifiers
 1..705
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-EB1-clf-i-07-0-UI"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="UI-R-EB1"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; UI-R-EB1 is a non-normalized Rat Duodenum library (RDU) constructed in pT73 PAC vector according to the procedure described by Bonaldo, Lennon & Soares (Genome Research 6: 791-806, 1996). The oligonucleotide used to prime first strand synthesis contained the sequence tag GTGGTTTCAT between the Not I cloning site and dT18 stretch. The Rat Duodenum tissue was provided by Tom Freeman of the Sanger Center.
 TAG_LIB=UI-R-EB1
 TAG_TISSUE=duodenum
 TAG_SEQ=GTGGTTTCAT
 BASE COUNT 226 a 136 c 165 g 177 t 1 others
 ORIGIN

Query Match 92.2%; Score 320.8; DB 13; Length 705;
 Best Local Similarity 99.1%; Pred. No. 9.8e-37;
 Matches 322; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 TGAACATCACTCTCTTGCAGAGTACCAACACTGTCATGTGATGATGCGGCCCTCT 60
 Db 326 TGAACATCACTCTCTTGCAGAGTACCAACACTGTCATGTGATGATGCGGCCCT 267
 QY 61 AGACCTCACCCAGCGGACACATCTTCCGTTACCTTGGGTCTGTGAGGTTCTGTCAAG 120
 Db 266 AGACCTCACCCAGCGGACACATCTTCCGTTACCTTGGGTCTGTGAGGTTCTGTCAAG 207
 QY 121 CGCTAGTGTAAACCGCGTCTGTACACCTAACTCACTGGCAAGACACAGTGTGGGCC 180
 Db 206 CGCTAGTGTAAACCGCGTCTGTACACCTAACTCACTGGCAAGACACAGTGTGGGCC 147
 QY 181 TTTCGACCACTAGAACAACTTTTTCATTTGACAGTGTGAGTGTGAGTGTGTTT 240
 Db 146 TTTCGACCACTAGAACAACTTTTTCATTTGACAGTGTGAGTGTGAGTGTGTTT 87
 QY 241 CATTTGATCTTTTGCATATGAGTGTAGCATATGTTTTCATGTATGATGACTTAATAATCCT 300
 Db 86 CATTTGATCTTTTGCATATGAGTGTAGCATATGTTTTCATGTATGATGACTTAATAATCCT 27
 QY 301 TGAATCATATAAAAAA 325

Db

RESULT 14
 AA899757/c
 LOCUS
 DEFINITION

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

REFERENCE
 AUTHORS
 TITLE

JOURNAL
 MEDLINE
 PUBMED
 COMMENT

On Apr 7, 1998 this sequence version replaced gi:3035111.
 Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: bento-soares@uiowa.edu
 CDNA Library Preparation: M. Ratima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics
 Seq primer: M13 Forward
 Location/Qualifiers
 1..465
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-E0-da-b-11-0-UI"
 /dev_stage="embryonic"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="UI-R-E0"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: NotI; Site 2: EcoRI; This library consists of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture."
 BASE COUNT 155 a 93 c 99 g 118 t
 ORIGIN

Query Match 91.4%; Score 318.2; DB 9; Length 465;
 Best Local Similarity 99.1%; Pred. No. 2.9e-36;
 Matches 320; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 TGAACATCACTCTCTTGCAGAGTACCAACACTGTCATGTGATGATGCGGCCCTCT 60
 Db 323 TGAACATCACTCTCTTGCAGAGTACCAACACTGTCATGTGATGATGCGGCCCTTT 264
 QY 61 AGACCTCACCCAGCGGACACATCTTCCGTTACCTTGGGTCTGTGAGGTTCTGTCAAG 120
 Db 263 AGACCTCACCCAGCGGACACATCTTCCGTTACCTTGGGTCTGTGAGGTTCTGTCAAG 204
 QY 121 CGCTAGTGTAAACCGCGTCTGTACAACTAACTCACTGGCAAGACACAGTGTGGGCC 180
 Db 203 CGCTAGTGTAAACCGCGTCTGTACAACTAACTCACTGGCAAGACACAGTGTGGGCC 144

|||||
 26 TGAATCATATAAAAAA

AA899757 465 bp mRNA linear EST 05-FEB-1999
 UI-R-E0-da-b-11-0-UI.s1 UI-R-E0 Rattus norvegicus cDNA Clone
 UI-R-E0-da-b-11-0-UI 3' similar to gi|464185|dbj|D25274|HUMPO2ST9
 Homo sapiens mRNA, clone: P02ST9, mRNA sequence.

AA899757
 AA899757.1 GI:4232251
 EST
 Rattus norvegicus (Norway rat)
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

1 (bases 1 to 465)
 Bonaldo, M.F., Lennon, G. and Soares, M.B.
 Normalization and subtraction: two approaches to facilitate gene discovery
 Genome Res. 6 (9), 791-806 (1996)

On Apr 7, 1998 this sequence version replaced gi:3035111.
 Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: bento-soares@uiowa.edu
 CDNA Library Preparation: M. Ratima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics
 Seq primer: M13 Forward
 Location/Qualifiers
 1..465
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-E0-da-b-11-0-UI"
 /dev_stage="embryonic"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="UI-R-E0"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: NotI; Site 2: EcoRI; This library consists of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture."
 BASE COUNT 155 a 93 c 99 g 118 t
 ORIGIN

Query Match 91.4%; Score 318.2; DB 9; Length 465;
 Best Local Similarity 99.1%; Pred. No. 2.9e-36;
 Matches 320; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 TGAACATCACTCTCTTGCAGAGTACCAACACTGTCATGTGATGATGCGGCCCTCT 60
 Db 323 TGAACATCACTCTCTTGCAGAGTACCAACACTGTCATGTGATGATGCGGCCCTTT 264
 QY 61 AGACCTCACCCAGCGGACACATCTTCCGTTACCTTGGGTCTGTGAGGTTCTGTCAAG 120
 Db 263 AGACCTCACCCAGCGGACACATCTTCCGTTACCTTGGGTCTGTGAGGTTCTGTCAAG 204
 QY 121 CGCTAGTGTAAACCGCGTCTGTACAACTAACTCACTGGCAAGACACAGTGTGGGCC 180
 Db 203 CGCTAGTGTAAACCGCGTCTGTACAACTAACTCACTGGCAAGACACAGTGTGGGCC 144

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 23, 2003, 10:43:52 ; Search time 2390.73 Seconds
(without alignments)
10338.975 Million cell updates/sec.

Title: US-09-717-321A-15

Perfect score: 1017

Sequence: 1 cccctatttgcagatt.....cccttggtctgtgaggttc 1017

Scoring table: IDENTITY_NDC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: em_estba.*

2: em_esthum.*

3: em_estin.*

4: em_estmu.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_estc.*

9: gb_esti.*

10: gb_est2.*

11: gb_est3.*

12: gb_est4.*

13: gb_est5.*

14: gb_est6.*

15: em_estfun.*

16: em_estom.*

17: em_gss_hum.*

18: em_gss_inv.*

19: em_gss_pln.*

20: em_gss_vrt.*

21: em_gss_fun.*

22: em_gss_mam.*

23: em_gss_mus.*

24: em_gss_pro.*

25: em_gss_rod.*

26: em_gss_phg.*

27: em_gss_vrl.*

28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	678.2	66.7	1796	11 AK081613	AK081613 Mus muscu
2	677.6	66.6	2269	11 AK034601	AK034601 Mus muscu
3	677.6	66.6	2270	11 AK047969	AK047969 Mus muscu
4	660.2	64.9	2308	11 AK076023	AK076023 Mus muscu

5	594.2	58.4	978	13	BUS14423	BUS14423
6	562	55.3	826	14	CA318577	CA318577
7	558.2	54.9	751	13	BQ177713	BQ177713
8	552.2	54.3	960	13	BUS11766	BUS11766
9	551.2	54.2	557	10	BG665146	BG665146
10	537.4	52.8	793	14	CA319273	CA319273
11	535.2	52.6	897	12	BI646446	BI646446
12	528.6	52.0	714	12	BI851489	BI851489
13	528.6	52.0	915	13	BQ931374	BQ931374
14	526	51.7	545	9	AA817948	AA817948
15	526	51.7	639	13	BQ201343	BQ201343
16	525.4	51.7	760	13	BQ191985	BQ191985
17	510.4	50.2	806	13	BQ444803	BQ444803
18	508.2	50.0	515	14	CD372599	CD372599
19	499.4	49.1	644	14	CA339407	CA339407
20	493.8	48.6	686	14	CB057853	CB057853
21	490.8	48.3	672	12	BI692509	BI692509
22	485	47.7	969	10	BF783603	BF783603
23	483.4	47.5	680	14	CD353147	CD353147
24	482.8	47.5	741	12	BI694684	BI694684
25	482.4	47.4	732	14	CB596051	CB596051
26	479	47.1	502	9	AI028951	AI028951
27	474.8	46.7	787	12	BG969782	BG969782
28	473.4	46.5	705	13	BQ200122	BQ200122
29	472.4	46.5	687	12	BM389059	BM389059
30	462.4	45.5	796	12	BI735385	BI735385
31	458.4	45.1	588	14	CA553516	CA553516
32	454.2	44.7	805	12	BI692090	BI692090
33	453	44.5	594	14	CB055806	CB055806
34	450.8	44.3	667	12	BQ042534	BQ042534
35	449.2	44.2	938	13	BQ28530	BQ28530
36	442.8	43.5	447	10	BG663683	BG663683
37	438.8	43.1	671	12	BM946306	BM946306
38	438.2	43.1	704	12	BM951935	BM951935
39	427.2	42.0	538	12	BM053597	BM053597
40	426.6	41.9	550	12	BM022481	BM022481
41	426.2	41.9	950	10	BF168374	BF168374
42	426	41.9	556	12	BI655954	BI655954
43	425.4	41.8	575	10	BF659623	BF659623
44	423.8	41.7	640	13	BQ780117	BQ780117
45	422	41.5	702	13	BQ176893	BQ176893

ALIGNMENTS

RESULT 1	AK081613	1796 bp	mRNA	linear	HTC 05-DEC-2002
LOCUS	AK081613	Mus musculus 16 days embryo head cDNA, RIKEN full-length enriched library, clone:Cl30051H24 product:RAS-related C3 botulinum substrate 1, full insert sequence.			
DEFINITION	AK081613	AK081613			
ACCESSION	AK081613	1 GI:26349264			
VERSION	AK081613	HTC; CAP trapper.			
KEYWORDS	HTC; CAP trapper.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	1	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
AUTHORS	Carninci, P. and Hayashizaki, Y.				
TITLE	High-efficiency full-length cDNA cloning				
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)				
MEDLINE	99279253				
PUBMED	10349636				
REFERENCE	2	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes			
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.				
TITLE	Normal normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)				
MEDLINE	20499374				
PUBMED	11042159				

REFERENCE AUTHORS	3	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000)
TITLE	20530913	
JOURNAL	11076851	
MEDLINE		
PUBMED		
REFERENCE AUTHORS	4	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G., Quackenbush, J., Schriml, L. M., Staubli, F., Suzuki, R., Tomita, M., Wagner, I., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S. and Hayashizaki, Y. Functional annotation of a full-length mouse cDNA collection Nature 409 (6821), 685-690 (2001)
TITLE	21085660	
JOURNAL	11217851	
MEDLINE		
PUBMED		
REFERENCE AUTHORS	5	The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002)
TITLE	6	(bases 1 to 1796)
JOURNAL		
MEDLINE		
PUBMED		
REFERENCE AUTHORS		Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission Submitted (16-Apr-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
TITLE		
JOURNAL		
MEDLINE		
PUBMED		
COMMENT		cDNA library was prepared and sequenced in Mouse Genome Encyclopedia project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL: http://genome.gsc.riken.go.jp/ URL: http://fantom.gsc.riken.go.jp/ Location/Qualifiers
FEATURES		

1.	1796	/organism="Mus musculus"
		/mol_type="mRNA"
		/strain="CS7BL/6J"
		/db_xref="PANTOM_DB:C130051H24"
		/clone="taxon:10090"
		/clone="C130051H24"
		/tissue_type="head"
		/clone_lib="RIKEN full-length enriched mouse cDNA library"
		/dev_stage="16 days embryo"
<1.	304	/note="unnamed protein product; RAS-related C3 botulinum substrate 1 (MGD GMI:97845, GB AK011072, evidence: BLASTN, 100%, match=987)
		putative"
		/codon_start=2
		/protein_id="BAC38272.1"
		/db_xref="GI:26349265"
		/translation="VRAXWYEVRRHHCNTPILVGTKLDLRDDKDTIEKLEKKLTP
		ITYPQGLAMAKEIGNVKYLECSALTQRLKTVDFDAIRAVLCPPPVKKERKRLLL"
		1781..1786
		/note="putative"
		1796
		/note="putative"
		polyA_signal
		polyA_site
		BASE COUNT 470 a 413 c 371 g 542 t
		ORIGIN
		Query Match 66.7%; Score 678.2; DB 11; Length 1796;
		Best Local Similarity 85.3%; Pred. No. 1.7e-183;
		Matches 871; Conservative 0; Mismatches 118; Indels 32; Gaps 9;
QY	1	CCCCATTCTTCTCAGATTAAAGTTCGCAAAATACCTTGTGAAGTTCGCGTTG 60
DB	548	CCCCATTCTTCTCAGATTAAAGTTCGCAAAATACCTTGTGAAGTTCGCGTTG 607
QY	61	CTGAGAACACGTAAGCACTTAAGTTCGCAAAATACCTTGTGAAGTTCGCGTTG 117
DB	608	CTGAGAACACGTAAGCACTTAAGTTCGCAAAATACCTTGTGAAGTTCGCGTTG 667
QY	118	TTCTGGGCTCAGGGTGCAGACCTCCCGTAGCTCCAGACGTTGACACAGCAGCC 177
DB	668	TTCTGGGCTCAGGGTGCAGACCTCCCGTAGCTCCAGACGTTGACACAGCAGCC 727
QY	178	TCCTTAATGACACGCTGCCATGTAAGCACTTAAGTTCGCAAAATACCTTGTGA 237
DB	728	TCCTTAATGACACGCTGCCATGTAAGCACTTAAGTTCGCAAAATACCTTGTGA 787
QY	238	AACTTTGTACTGTACGTAC---GATGGGTGAACAGCTCTCTCTTGAATTCATAG 294
DB	788	AACTTTGTACTGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGT 847
QY	295	AGTTCTTAAATATACAGCTGACGGCTTCTGAGGCTTGAACAGACCTGCGTCTCG 354
DB	848	AGTTCTTAAATATACAGCTGACGGCTTCTGAGGCTTGAACAGACCTGCGTCTCG 907
QY	355	TCGTGCTCTAAGCAAGTATTCCTGCTAGTCTGCTGCTGCTGCTGCTGCTGCTGA 414
DB	908	TCGTGCTCTAAGCAAGTATTCCTGCTAGTCTGCTGCTGCTGCTGCTGCTGCTGA 965
QY	415	ACACGACGTCATCAAGGAGACAGACAGTATTTGCTAATATGAAGTAGAGATTAT 474
DB	966	ACACGACGTCATCAAGGAGACAGACAGTATTTGCTAATATGAAGTAGAGATTAT 1025
QY	475	ACATACATTTACATGAGTATTCACACTGAATAAAGTGCACGGTAAAGCTTTTA 534
DB	1026	ACATACATTTACATGAGTATTCACACTGAATAAAGTGCACGGTAAAGCTTTTA 1085
QY	535	ACGTTAATTTCTGTCACACAGTAGATGACAAATGCCGATCTTATCAGTGTCTC 589
DB	1086	ACGTTAATTTCTGTCACACAGTAGATGACAAATGCCGATCTTATCAGTGTCTC 1145
QY	590	-----TCTTGAGCCCCCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 642


```
putative"
/codon_start=1
/best_local_similarity 85.3%; Score 677.6; DB 11; Length 2269;
/db_xref="GI:26330057"
/translat="MQAIKCVVGVGAVGKTCCLISYTNAPPGYIPIVFDNYSANV
MVDKPNVLGMDTAGOEDYDLRLPRLPOTDVLICPLSLVSPASFENVRKMYPEVR
HHCENPILVGTIKLIDRDDKDTIEKLEKILPTITYPQGLAMAKEIGAVKYLECSAL
TORGLKTVDFDEAIRAVLCPPVPPVKKRKKCLLL"
polyA_signal
2252..2257
/note="putative"
polyA_site
2269
/note="putative"
/number="putative"
BASE COUNT 555 a 569 c 502 g 643 t
ORIGIN
Query Match 66.6%; Score 677.6; DB 11; Length 2269;
Best Local Similarity 85.3%; Pred. No. 2.9e-183;
Matches 870; Conservative 0; Mismatches 119; Indels 31; Gaps 9;
QY 1 CCCCTATTCTCTCAGATTAGAAATGCCCCAAATACCTTGTGAAGTTCGCTTGGTGG 60
Db 1020 CCCCCTATTCTCTCAGATTAGAAATGCCCCAAATACCTTGTGAAGTTCGCTTGGTGG 1079
QY 61 CTGGAACACCTAAGCACTAAGCTGTGTGAGAGACCTTTGTCCTTAAGAACGCTGCAGC 117
Db 1080 CTGGAACACCTAAGCACTAAGCACTTCTTTGAGAGACTTCTGTGTAAGAACGCGCAGC 1139
QY 118 TTCTGGGCTCAGGGTGCAGACCCCTCCGCTAGCTCCAGACCGTGTGCACAGCAGCAGC 177
Db 1140 TTCTGGAGCCAGGGTGCAGACCCCTCCAGAGTTCAGACCGTGTGCACAGCAGCAGC 1199
QY 178 TCCTTAATGACACGCTGCCATGTAAAGCACTGTAACTTATCAGCCCATGCTCATACGT 237
Db 1200 TCCTTAATGACACGCTGCCATGTAAAGCACTGTAACTTATCAGCCCATGCTCATACGT 1259
QY 238 AACTTTGACTGTAGTGCAC---GATGGGTGTAAACAGCTCTGCTCTTTGATTTTCAATG 294
Db 1260 AACTTTGACTGTAGTGCACAGTGTAGTGTGACAGCTCAGCTCTTTGATTTTCAATG 1319
QY 295 AGTTCTCTAAATATACAGCTGACCGGCTCTCTGAGGCTTTGAACAGAACTCTGGCTCCTG 354
Db 1320 AGTTCTCTAAATATACAGCTGACCGGCTCTCTGAGAACTTTGAACAGAACTCTGGTCTCTG 1379
QY 355 TGTGCTCTTAAGAAAGTATCTGTTCTTCTAGTCTGGGTGTGCTGGGTGAGTGTGTA 414
Db 1380 TGTGCTCTTAAGAAAGTATCTGTTCTTCTAGTGTGGGTGTGCTGGGTGGA--GTGTGAA 1437
QY 415 ACACGACGTCATCAAGAGGAGACAGACGATTTTGTACTAATATGAAGTAGAGATTAAAT 474
Db 1438 ACACGACATGATCAAGAGGAGACAGACGATTTTGTACTAATATGAAGTAGAGATTAAAT 1497
QY 475 ACACATCATTTGATGAGTGAATCACTGAATAAAGTGTACCGGTAAAGCTTTTGA 534
Db 1498 ACACATCATTTGATGAGTGAATCACTGAATAAAGTGTACCGGTAAAGCTTTTGA 1557
QY 535 ACGTTAAATTTCTCTCAACAGTAGATGACAAATGGCCGATCTTATCAGTGTCTC--- 589
Db 1558 ACGTTAAATTTCTCTCAACAGTAGATGACAAATGGCCGATCTTATCAGTGTCTCTTGAG 1617
QY 590 -----TCTTGAGCCCTCTCCCTGCTGCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 642
Db 1618 CATCCCTTGGCGCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 1677
QY 643 TATTAAATGCGCATCTCTCAGTGTGCTAATCACTGAATAAAGTGTACCGGTAAAGCTTTTGA 702
Db 1678 TATTAAATGCGCATCTCTCAGTGTGCTAATCACTGAATAAAGTGTACCGGTAAAGCTTTTGA 1735
QY 703 TTCTTAACGACGATATGCTGACCTGTGCTAATCACTGAATAAAGTGTACCGGTAAAGCTTTTGA 757
Db 1736 TTCTTAACGACGATATGCTGACCTGTGCTAATCACTGAATAAAGTGTACCGGTAAAGCTTTTGA 1795
QY 758 CGGAGATTTTTTGGTAGATAGTAGAGTGCCTTCTGTTTTCACCTTCTTTTACTCAG- 816
Db 758 CGGAGATTTTTTGGTAGATAGTAGAGTGCCTTCTGTTTTCACCTTCTTTTACTCAG- 816
```

1796 TTTGTTTTTTTGGTAGTAGAAGTGGTCTCTGTTTTCGGCTTACTTTACTCAGA 1855

817 CTGACTAGTCTTCCCTTCGTTTCTAGTAAGTGGTGTAGAAATCACTGCTGGGCTT 876

1856 GCTAGTGTAGTCTTCCCTTCGTTTCTAGTAAGTGGTGTAGAAATCACTGCTGGGCTT 1915

877 TACAGTGTAAACATTTTATGATATTCTGAAACATCACTGTTTGGCCAGAGTACCAACA 936

1916 TACGTTTAAACATTTTATGATATTCTTAAACATGTAACCTTCTAAGACAGTACCAACA 1975

937 CTGTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 996

1976 CTGTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2032

AK047969 2270 bp mRNA linear HFC 05-DEC-2002
Mus musculus 16 days embryo head cDNA, RIKEN full-length enriched
library, clone: C130025F01 product: RAS-related C3 botulinum
substrate 1, full insert sequence.

AK047969 GI:26339063
HTC; CAP trapper.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636

2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159

3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, K.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861

4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Arai, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G.,
Quackenbush, J., Schriml, L. M., Staubli, P., Suzuki, R., Tomita, M.,
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C.,
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,
Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,
Rang, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,
Toyo-oka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L.,

Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
and Hayashizaki,Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
21085660
MEDLINE
PUBMED
11217851
REFERENCE
AUTHORS

The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 2270)

Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
Okazaki,Y., Saiki,D., Saichou,H., Sakai,C., Sakai,K., Sakazume,N.,
Sano,H., Sasaki,T., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
Direct Submission

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL:http://genome.gsc.riken.go.jp/
URL:http://fantom.gsc.riken.go.jp/
Location/Qualifiers

FEATURES

source
1..2270
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:C130025F01"
/db_xref="taxon:10090"
/clone="C130025F01"
/tissue_type="head"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="16 days embryo"
200..778

CDS

/note="unnamed protein product; PAS-related C3 botulinum
substrate 1 (MGI:97845, GB|AK011072, evidence: BLASTN,
100%, match=987)
putative"
/codon_start=1
/protein_id="BAC33203.1"
/db_xref="GI:26339064"
/translation="MQAICVGVGVGAVGKTCLLISYTNAPFGEYIPVFNYSNV
WVGRPNVGLMDTQEDYDLPLSPQDVELICFSLVSPAFENVRKAYEVR
HPCNPPIILVGTKLDRDKDTIEKLEKKLPIIYPQGLAWAKEIGAVKYLEUSAL
TORGLKTVDEAIRAVLPPPPVKKRRKKLLI"
2254..2259
/note="putative"
2270
/note="putative"

polyA_signal

polyA_site

BASE COUNT 554 a 569 c 503 g 644 t
ORIGIN

Query Match

Best Local Similarity 66.6%; Score 677.6; DB 11; Length 2270;
85.3%; Pred. No. 2.9e-183;

	Matches	870;	Conservative	0;	Mismatches	119;	Indels	31;	Gaps	9;
QY	1	CCCCATTTCTTCTGCTCAGATTAAAGAAATTC	CCCAAAATACCTTGTGA	AACTAAGTTGCGTTG	60					
Db	1022	CCCCATTTCTTCTGCTCAGATTAA	AAATTC	CCCTTCTGAACTAAGTTGCGTTG	1081					
QY	61	CTGAGAACCTAAGCTAAGCTGTTGAGAGA	---	CTTGTCTTAAAGACATGCGAC	117					
Db	1082	CTGAGAACCTAAGCTAAGCTTCTTGAGAG	AACTTCTTGAGAGACTTCTGTTGTA	AGACCGGACG	1141					
QY	118	TTCTGGGCTCAGGGGTGCAGACCTCCGTA	CTCCAGTCCGAGACCGTGTGACAGACAC	CGCC	177					
Db	1142	TTCTGGAGCCAGGGGTGCAGACCTTCC	AGAGTTC	CCAGACCGTGTGACAGACACG	1201					
QY	178	TCCTTAATGACACCTGCCATGTAAACG	CACTGTAACTTATCAGCCCATGCTCAT	AGT	237					
Db	1202	TCCTTAATGACACCTGCCATGTAAACG	CACTGTAACTTATCAGCCCATGCTCAT	AGT	1261					
QY	238	AACCTTCTAAGTACGTCAC	---GATGGGTGAACAGCTCTGCTCTTTG	ATTTTCAATAGT	294					
Db	1262	AACCTTCTAAGTACGTCACAGT	GATGAGTGTGACAGCTCAGCTCCTTTG	ATTTTCAATAGT	1321					
QY	295	AGTTCTCTAAATACCACTGACCGGCTT	CGAGGCTTTGAAACAGAACTCTGCTCCT	GT	354					
Db	1322	AGTTCTCTAAAGACCCGCGACTAGCT	TTTGACAGACTTTGAACAGAACTCTGCTCCT	GT	1381					
QY	355	TGTTGCCCTTAAACGAAGTATTTCTT	CTTAGTCTGGTGTGCTGGTGGAGTGTGTGA	A	414					
Db	1382	TGTTGCCCTTAAAGTATTTCTGTT	CTTAGTGTGGTGTGCTGGTGGAGTGTGTGA	A	1439					
QY	415	ACAGAGCTCATCAAGGAGACAGACAG	TATTTGACTAATATGAAGTAGAGATTAA	TTT	474					
Db	1440	ACAGACATGATCAAGGAGACAGACAG	TATTTGACTAATATGAAGTAGAGATTAA	TTT	1499					
QY	475	ACACTAATGTACATGAGTAAATTC	AACTGAATAAAGTGTACGGGTAAAGCTTT	TTA	534					
Db	1500	ACACTAATGTACATGAGTAAATTC	AACTGAATAAAGTGTACGGGTAAAGCTTT	TTA	1559					
QY	535	ACGTTAAATTTCTGTCAAACAGTAG	ATGACAAATGGCCGATCTTATCAGTGTCTC	---	589					
Db	1560	ACGTTAAATTTCTGTCANACAGTAG	GTGACAAATGGCCGATCTTATCAGTGTCTC	TTGAG	1619					
QY	590	-----TC	TGAGCCCCCTTCCCCTGCTGTCCCTCCC	CAGATGGGGTGTGAGTCCA	642					
Db	1620	CATCCCTTCGCGCCCCCAACCTCC	CACTGCCGTCCCTCCCAGATGGGGTGTGAGTCCA	1679						
QY	643	TATTTAACTGGCCATCCTCACAG	TTGCTAACTTAGCAAGTGTCTTTTAGACCC	CCCC	702					
Db	1680	TATTTAACTGGCCATCCTCACAG	TTGCTAACTTAGCAAGTGTCTTTTAGACCC	CCCC	1737					
QY	703	TTCTTAACGAGCAATATGTGACCT	GTACTATAAAGATCTTCTGTATATGCA	----	757					
Db	1738	TTCTTAACGAGCAATATGTGACCT	GTACTATAAAGATCTTCTGTATACGCA	GATTT	1797					
QY	758	CGGAGATTTTTTTGGTAGATAG	TAGAAAGTCGGTCTCTGTTTCCACCTT	CTTACTCAG	816					
Db	1798	TTTGTTTTTTTTGGTAGCTCAG	TAGAGTGGCTCTGTTTTCGCTTACTT	TACTCAGA	1857					
QY	817	CTGACTAGTGTCTCCCTTCGTTT	CTTAGTAACTGGGTGTAGAAATCAGT	CTGCGGCTT	876					
Db	1858	GCTAGTGTGCTTCCCTTCGTTT	CTTAGTAACTGGGTGTAGAAATCAGT	CTGCGGCTT	1917					
QY	877	TACAGTTTTTAAACATTTT	TAGATTTCTGAAACATCACTGCTT	GTGCCAGAGTACCACA	936					
Db	1918	TACGGTTTTTAAACATTTT	TAGATTTCTTAACTATGAACTTCT	TAACAGTACCACA	1977					
QY	937	CTGTCAATGTGATGTGCG	CCCCCTCTAGACCTCACCCACGCGGAC	CATGCTTCCGCT	996					
Db	1978	CTGTCAATGTGATGTGCTG	-----CCTCTAGACCTCGCCCACT	GTGACAGACGACCTTCTCT	2034					

RESULT 4
AK076023

LOCUS	AKO76023	2308 bp	mRNA	linear	HTC 07-DEC-2002
DEFINITION	Mus musculus 10 days embryo whole body cDNA, RIKEN full-length enriched library, clone:2610100016 product:RAS-related C3 botulinum substrate 1, full insert sequence.				
ACCESSION	AKO76023				
VERSION	AKO76023.1 GI:26344957				
KEYWORDS	HTC; CAP trapper.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Sasaki, N., Carninci, P., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.				
TITLE	High-efficiency full-length cDNA cloning				
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)				
MEDLINE	99279253				
PUBMED	10349636				
REFERENCE	2				
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Sasaki, N., Carninci, P., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.				
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)				
MEDLINE	20493374				
PUBMED	11042159				
REFERENCE	3				
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.				
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer				
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)				
MEDLINE	20530913				
PUBMED	11076861				
REFERENCE	4				
AUTHORS	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaudo, I., Pesole, G., Quackenbush, J., Schriml, L. M., Stauber, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohsaki, S. and Hayashizaki, Y.				
TITLE	Functional annotation of a full-length mouse cDNA collection				
JOURNAL	Nature 409 (6821), 685-690 (2001)				
MEDLINE	21085660				
PUBMED	11217851				
REFERENCE	5				
AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.				
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs				
JOURNAL	Nature 420, 563-573 (2002)				
REFERENCE	6 (bases 1 to 2308)				
AUTHORS	Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,				
	Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numata, C., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yamanaka, I., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.				
	Direct Submission				
	Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216]				
	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.				
	Please visit our web site for further details.				
	URL: http://genome.gsc.riken.go.jp/				
	URL: http://fantom.gsc.riken.go.jp/.				
	Location/Qualifiers				
FEATURES	1..2308				
SOURCE	/organism="Mus musculus"				
	/mol_type="mRNA"				
	/strain="CS7BL/6J"				
	/db_xref="FANTOM_DB:2610100016"				
	/db_xref="MGI:1898616"				
	/db_xref="taxon:10090"				
	/clone="2610100016"				
	/tissue type="whole body"				
	/clone_lib="RIKEN full-length enriched mouse cDNA library"				
	/dev_stage="10 days embryo"				
	239..817				
CDS	/note="unnamed protein product; RAS-related C3 botulinum substrate 1 (MGD MGI:97845, GI AK011072, evidence: BLASTN, 100%, match=987)"				
	putative				
	/codon_start=1				
	/protein_id="BAC36128.1"				
	/db_xref="GI:26344958"				
	/db_xref="MGI:97845"				
	/translation="MQAIKCVVGVGAGVKTCLLISYTNAPFGEYIPVFDNYSANVWDGKPVNLGLWDTAGQEDYDLRLPLSPQTDVFLICSLVSAASPVRAKWKYPEVRHCPNPILVGTGLDRLDDDDDTIEKLEKLTPTTPYQGLAWAKEIGAVKYLECSALTGRLKTVFEAIRAVLCPPPVKKRRKCLLL"				
	2293..2298				
	/note="putative"				
	2308				
	/note="putative"				
polyA_site	559 a 580 c 518 g 650 t 1 others				
polyA_site	2308				
BASE COUNT	559 a 580 c 518 g 650 t 1 others				
ORIGIN	Query Match 64.9%; Score 660.2; DB 11; Length 2308;				
	Best Local Similarity 84.8%; Pred. No. 3e-178;				
	Matches 866; Conservative 0; Mismatches 123; Indels 32; Gaps 10;				
QY	1 CCCTATTCTGCTCAGATTAGAAATGCCAAATACCTTGCACTAGTTCGGTGTG 60				
Db	1061 CCCCCATTCTGCTCAGATTAAAAATGGCAAATACCTTCTGAACTAGTTCGGTGTG 1120				
QY	61 CTGAGACACGTAAGCACTAAGCTGTGTGAGAGA--CTTTGTCTTAAAGAACTGCAGC 117				
Db	1121 CTGAGAACCTTAAGCACTAACTCTCTTGAGAGACTTCTGTTGCTAAGAGACCGCAGC 1180				
QY	118 TTCTGG-GGTCAGGGGTGCAGACCCCTCCCGTAGCTCCAGACCGGTGTGCACACGAC 176				
Db	1181 TTCTGGAGCCAGGGGTGCAGACCCCTCCAGAGTCCAGACCGGTGTGCACACGAC 1240				
QY	177 CTCCTTAATGACAGCGTGCCATGTAAGCGACCTGTACTTATCAGCCCATGCTATTACG 236				

```
Db 1241 CTCCTTAATGACAGCTTGCCATGTAAACGACCTGTAACTATACGCCCATGTTCTATTAG 1300
QY 237 TAACCTTTGACTGTACGTAC---GATGGGTGAACAGCTCTGCTCTTTTGATTTCAATAGT 293
Db 1301 TAACCTTTGACTGTACGTACAGTGTACAGTGTACAGCTCAGCTCTTTTGATTTCAATAGT 1360
QY 294 GAGTTCTCTAAATACACAGCTGACCGGCTCTCGAGGCTTTGAACAGAACTCTGCTCTCT 353
Db 1361 GAGTTCTCTAAAGACACCGGCTGACTGCTTTTGCAGACTTTTGAACAGAACTCTGCTCTCT 1420
QY 354 GTGTTGCTCTTAAGGAAGTATTTCTTCTAGTCTGGGTGTGCTGGGTGAGTGTGGA 413
Db 1421 GTGTTGCTCTTAAGGAAGTATTTCTTCTAGTCTGGGTGTGCTGGGTGGA--GTGGA 1478
QY 414 AACAGGAGCTCATCAAGAGGACAGACAGTATTTTGACTATATGAAGTAGAGTTAAT 473
Db 1479 AACAGGAGCTCATCAAGAGGACAGACAGTATTTTGACTATATGAAGTAGAGTTAAT 1538
QY 474 TACACTACATTTGATGAGTAACTCACTGAATGAATGAAGTGTACCGGTAAAGCTTTT 533
Db 1539 TACACTACATTTGATGAGTAACTCACTGAATGAATGAAGTGTACCGGTAAAGCTTTT 1598
QY 534 AACGGTTAAATTTCTGTCACACAGTAGATGACAAATGGCCGATCTTATCAAGTGTCTC--- 589
Db 1599 AACGGTTAAATTTCTGTCACACAGTAGTGAACAATGGCCGATCTTATCAAGTGTCTTGA 1658
QY 590 -----TCTTGAGCCCTCCCTCCCTGCTGCTCCCTCCAGATGGGGGCTTGAGTCC 641
Db 1659 GCATCCCTTCGCGGCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 1718
QY 642 ATATTTAACTGGCCATCTCACAGTGTCTAACTAGCAAGTGTCTTTCTTTAGACCCC 701
Db 1719 ATATTTAACTGGCCATCTCACAGTGTCTAACTAGCAAGTGTCTTTCTTTTGA-CAAACC 1776
QY 702 CTTCTTAACGAGCAATATGCTGACCTGTACTATAAGATCTTTCTGATAATGCA-----T 756
Db 1777 CTTCTTAACGAGCAATATGCTGACCTGTACTATAAGATCTTTCTGATAAGCAAGAT 1836
QY 757 TCGAGAGATTTTGGTAGATAGTAGAAGTGGTCTGTTTCACTTCTCTTTACTCAG 816
Db 1837 TTTTGTGTTTTTGGTAGTCTAGTAGAAGTGGTCTGTTTCACTTCTCTTTACTCAG 1896
QY 817 -CTGACTAGTCTCCCTGCTGTTTCTAGTAACTGGGTGTAGAAATCACTGCTCGGCT 875
Db 1897 AGTAGTAGTCTCTCTGTTTCTAGCAACTAGTGTGTGCAAAATCATGTGTTGACGT 1956
QY 876 TTACAGTTTTTAACTATTTAGATATTTCTGAACATCACTGTCTGCGAGAGTACCAAC 935
Db 1957 TTACGGTTTTTAACTATTTAGATATTTCTTAACATATGAACCTTTAAACAGTACCAAC 2016
QY 936 ACTGTCTAGTGTATGATGCGGCCCTCTAGACCTCAACCGCGGACACATGTTCCGG 995
Db 2017 ACTGTCTAGTGTATGATGCGGCCCTCTAGACCTCAACCGCGGACACATGTTCCGG 2073
QY 996 T 996
Db 2074 T 2074

RESULT 5
BU514423
LOCUS
DEFINITION BU514423 978 bp mRNA linear EST 12-SEP-2002
AGENCOURT 10119699 NIH MGC 134 Mus musculus cDNA clone
IMAGE:6511045 5', mRNA sequence.
ACCESSION BU514423
VERSION BU514423.1 GI:22821949
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
REFERENCE 1 (bases 1 to 978)
```

AUTHORS
TITLE
JOURNAL
COMMENT

NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. David Rowe
cDNA Library Preparation: Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLN)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ILLN at:
<http://image.llnl.gov>
Plate: LLN14080 row: o column: 14
High quality sequence start: 17
High quality sequence stop: 714.
Location/Qualifiers
1. 978
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6511045"
/tissue_type="undifferentiated limb"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_134"
/note="vector: pCMV-SPORT6.1.cdb; Site 1: EcoRV; Site 2:
NotI; Cloned unidirectionally. Primer: Oligo dt. Average
insert size 1.7 kb. Constructed by ResGen, Invitrogen
Corp. Note: this is a NIH MGC Library."
BASE COUNT 249 a 224 c 201 g 301 t 3 others
ORIGIN

FEATURES
source

Query Match 58.4%; Score 594.2; DB 13; Length 978;
Best Local Similarity 85.8%; Pred. No. 1.9e-159;
Matches 762; Conservative 0; Mismatches 94; Indels 32; Gaps 8;
QY 1 CCCCTATCTGCTCAGATTAAGTAATGCCAAATACCTTGTGAAGTGGTGTG 60
Db 70 CCCCATCTTGTCTCAGATTAAGTAATGCCAAATACCTTGTGAAGTGGTGTG 129
QY 61 CTGAGAACAGCTAAGCACTAAGCTGTGTGAGAGA---CTTTGTCTTAAAGAGCTGCAGC 117
Db 130 CTGAGAACACTAAGCACTAAGCTGTGTGAGAGACTTCTGTCTAAGAACCGGAGC 189
QY 118 TCTGGGCTCAGGGGTGACAGACCTCCCTGCTAGCTCCAGACCGGTGTGACACAGCAGCC 177
Db 190 TCTGGAGCAGGGGTGACAGACCTCCCTGCTAGCTCCAGACCGGTGTGACACAGCAGCC 249
QY 178 TCCTTAATGACAGCTGCCATGTAAACCACTGTAACTTATCAGCCCATGCTCATAGT 237
Db 250 TCCTTAATGACAGCTGCCATGTAAACCACTGTAACTTATCAGCCCATGCTCATAGT 309
QY 238 AACTTTGTACTGTACGTACAC---GATGGGTGTAAACAGCTCTGCTTTGATTTCAATAGT 294
Db 310 AACTTTGTACTGTACGTACAGTGTGAGTGTGACAGCTCAGCTCTTTGATTTCAATAGT 369
QY 295 AGTTCTCTAAATACAGCTGACCGGCTTCTGAGGCTTTTGAACAGAACTCTGGCTCCTG 354
Db 370 AGTTCTCTAAAGACACCGGCTAGCTTTTGAGAGCTTTTGAACAGAACTCTGGTCTCTG 429
QY 355 TGTGCTCTCAAGAGTATTTCTGCTAGTGTGGGTGTGCTGGGTGAGTGTGGA 414
Db 430 TGTGCTCTCAAGAGTATTTCTGCTAGTGTGGGTGTGCTGGGTGAGTGTGGA--GTGTGA 487
QY 415 ACAGCAGCTCATCAAGAGGACAGACAGTATTTTGACTATATGAAGTAGAGTAAATTT 474
Db 488 ACAGCAGTGTCAAGAGGACAGACAGTATTTTGACTATATGAAGTAGAGTAAATTT 547
QY 475 ACACACTATGTACATGGAGTAAATTCACACTGAATAAAGTGTCAACGGTAAAGCTTTTA 534
Db 548 ACACACTATGTACATGGAGTAAATTCACACTGAATAAAGTGTCAACGGTAAAGCTTTTA 607
QY 535 ACGTTAATTTCTGTCAACAGTAGATGCAAAATGGCCGATCTTATCAGTGTCTC----- 589

Db 608 ACGGTAAATTTCTGTCACAAACAGTAGTGAGCAAAATGCCGATCTTATCAGTGTCTCTTGAG 667
 QY 590 -----TCTTGAGCCCCCTTCCCTCTGCTCCCTCCAGATGGGCGTGTGATCCA 642
 Db 668 CATCCCTTCGCGCCCCCACCCTCCACTCCCTCCCTCCAGATGGGCGTGTGATCCA 727
 QY 643 TATTTAAACTGCGCATCTCCACAGTTCCTAACTTAGCAAGTGTCTTTCTTTAGGACCCCC 702
 Db 728 TATTTAAACTGCGCATCTCCACAGTTCCTAACTTAGCAAGTGTCTTTCTTTAG--ACCCC 785
 QY 703 TTTCTAACGAGCAATATGCTGACCTGTACTATAAGATCTTTCTGATAAANGCA-----T 756
 Db 786 TTTCTTACGAGCAATATGCTGACCTGTACTATAAGATCTTTCTGATAAANGCAATTT 845
 QY 757 TCGGAGATTTTTTGTAGAT--AGTAGAGTGGCTCTCTGTTTCTTCACTTCTTACTC 814
 Db 846 TTTGTTTTTTTTTGTAGTCAAGTAGAGTGGCTCTCTGTTTCTTCACTTCTTACTC 905
 QY 815 --AGCTGACTAGTGTCTCCCTTCTGTTTCTAGTAACTGGGTGTAGAAA 860
 Db 906 AGAGCTAGTGTGGTCTCTTCTAGTAACTGGGTGTAGAAA 953

RESULT 6

CA318577

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CA318577 826 bp mRNA linear EST 26-NOV-2002
 UI-M-FWO-cbr-0-14-0-UI.r1 NIH_BMAP_FWO Mus musculus cDNA clone
 IMAGE: 6813639 5', mRNA sequence.
 CA318577
 CA318577.1 GI:24536701
 EST.
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NIH-MGC http://img.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cga@rmail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library preparation by: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)

Seq primer: pYX-5.

Location/Qualifiers

1..826

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6"

/db_xref="taxon:10090"

/clone="IMAGE: 6813639"

/tissue_type="whole brain"

/dev_stage="embryo 13.5,14.5,16.5,17.5dpc"

/lab_host="PH10B (T1 phage resistant)"

/clone_lib="NIH_BMAP_FWO"

/note="Organ: Brain; Vector: pYX-Asc; Site: 1; EcoR I;

Site 2: Not I; The library was constructed according
 Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
 1996. Denatured RNA was size fractionated on a 1% agarose
 gel. First strand cDNA synthesis was primed with oligo-dT
 primer containing a Not I site. Double strand cDNA was
 size selected according to mRNA size fraction, ligated
 with EcoR I adaptor, digested with NotI and then cloned
 directionally into pYX-Asc vector. The library tag
 sequence located between the Not I site and the polyA tail
 is AGCGAGACAG. This library was created for the University

FEATURES

source

Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
 Developing Mouse Nervous System', supported by National
 Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
 program coordinator." 242 t 1 others

BASE COUNT 226 a 195 c 162 g 242 t 1 others

ORIGIN

Query Match 55.3%; Score 562; DB 14; Length 826;

Best Local Similarity 89.7%; Pred. No. 3.1e-150;

Matches 677; Conservative 0; Mismatches 56; Indels 22; Gaps 6;

QY 1 CCCCTATTCTTCTCAGATTAAAGAAATGCCAAATACCTTGTCAATAAGTTCGTTGTG 60
 Db 76 CCCCCATTTCTGCTCAGATTAAAAATGCCAAATACCTTGTCAATAAGTTCGTTGTG 135
 QY 61 CTGAGAACAGTAAGCACTAACTGTTGAGAGA---CTTTGTCCCTTAAAGAACTCCAGC 117
 Db 136 CTGAGAACACCTTAAGCACTAACTCTTTGAGAGACTTCTTTGCTTAAGAACCCGAGC 195
 QY 118 TTCTGGGCTCAGGGGTGCAGACCCCTCCCTAGTCCAGACCGTGTGACACAGCAGCC 177
 Db 196 TTCTGGAGCCAGGGGTGCAGACCCCTCCAGAGTTCCAGACCGTGTGACACAGCAGCC 255
 QY 178 TCCTTAATGACACGCTGCCATGTAACGACCTGTAACTTATCAGCCCATGCTCATTCAGT 237
 Db 256 TCCTTAATGACACGCTGCCATGTAACGACCTGTAACTTATCAGCCCATGCTCATTCAGT 315
 QY 238 AACTTTGTAAGTACGTAC--GATGGGTGTAACAGCTCTGCTCTTCAATTCATGATG 294
 Db 316 AACTTTGTAAGTACGTACGTACGTGAGTGTGACAGCTCAGCTCTTTGATTTCAATG 375
 QY 295 AGTTCTCTAAATACCAAGTACCGGCTTCTGACAGCTTTGAACAGAACTCTGGCTCCTG 354
 Db 376 AGTTCTCTAAAGACCAAGCGACTAGCTTTTGCAGACTTTGAACAGAACTCTGGTCCG 435
 QY 355 TGTGCTCTTAACGAAGTATTCGTTCTAGTCTGGTGTGCTGGTGGAGTGTGTGAA 414
 Db 436 TGTGCTCTTAATGAAGTATTCGTTCTAGTGTGGTGTGCTGGTGTGGA--GTGTGAA 493
 QY 415 ACAGACGCTCATAAGAGGAGACAGACAGTATTTGACTAATATGAAGTAGAGATTATTT 474
 Db 494 ACAGACATGATCAAGAGGAGACAGACAGTATTTGACTAATATGAAGTAGAGATTATTT 553
 QY 475 ACACATACATGTACATGGAGTAATTCAACTGAATAAAAGTGTACGGGTAAAGCTTTTA 534
 Db 554 ACACATACATGTACATGGAGTAATTCAACTGAATAAAAGTGTACGGGTAAAGCTTTTA 613
 QY 535 ACGTTAATTTCTGTCAAAAGTATGACAAATGCCCGATCTTATCAGTGTCTC----- 589
 Db 614 ACGTTAATTTCTGTCAAAAGTATGAGTGTGACAAATGCCCGATCTTATCAGTGTCTCTGAG 673
 QY 590 -----TCTTGAGCCCCCTTCCCTCTGCTCCCTCCACAGATGGGCGTTGAGTCCA 642
 Db 674 CATCTCTTCGCGCCCCCACCCTCCACTGCGCTCCCTCCACAGATGGGCGTTGAGTCCA 733
 QY 643 TATTTAACTGGCCATCTCCTCACAGTTGCTAACTTAGCAAGTGTCTTTCTTTAGACCCCC 702
 Db 734 TATTTAACTGGCCATCTCCTCACAGTT-CTAACTTAGCAAGTGTCTTTCTTTA-GAACCCC 791
 QY 703 TTTCTAACGAGCAATATGCTGACCTGTACTATAAGTGTGACTATAA 737
 Db 792 TTTCTAACGAGCAATATGCTGAGTGTACTATAA 826

RESULT 7

BQ177713

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

BQ177713 751 bp mRNA linear EST 30-APR-2002
 UI-M-ERO-bwp-j-03-0-UI.r1 NIH_BMAP_ERO Mus musculus cDNA clone
 IMAGE: 5700194 5', mRNA sequence.
 BQ177713
 BQ177713.1 GI:20353205
 EST.
 Mus musculus (house mouse)

ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 751)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Dr. James Lin, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)
 Seq primer: pYX-5.
 Location/Qualifiers
 1. 751
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:5700194"
 /tissue_type="whole brain"
 /dev_stage="embryo 15.5 dpc"
 /lab_host="DH10B (TI phage resistant)"
 /clone_lib="NIH BMAP ERO"
 /note="Organ: brain; Vector: pYX-Asc; Site 1: EcoR I;
 Site 2: Not I; The library was constructed according to
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
 1996. Denatured mRNA was size fractionated on a 1% agarose
 gel. First strand cDNA synthesis was primed with an
 oligo-dT primer containing a Not I site. Double stranded
 cDNA was size selected according to mRNA size fraction,
 ligated with EcoR I adaptor, digested with Not I, and then
 cloned directionally into pYX-Asc vector. The library tag
 sequence located between the Not I site and the polyA tail
 is GTCGCTGGAA. This library was created for the
 University of Iowa Mouse Brain Molecular Anatomy Project
 (BMAP). 'Gene Discovery in the Developing Mouse Nervous
 System', supported by National Institutes of Mental Health
 (NIMH), Hemin Chin, Ph.D., program coordinator."
 198 a 180 c 154 g 217 t 2 others

BASE COUNT 198 a 180 c 154 g 217 t 2 others

ORIGIN

Query Match 54.9%; Score 558.2; DB 13; Length 751;
 Best Local Similarity 89.6%; Pred. No. 3.7e-149;
 Matches 673; Conservative 0; Mismatches 55; Indels 23; Gaps 6;

QY 1 CCCCTATTCCTGCTAGATTAGAAATGCGAAATACCTTGTGAACCTAAGTTCGTTGTG 60
 |||||
 Db 6 CCCCCATTCCTGCTAGATTAGAAATGCGAAATACCTTGTGAACCTAAGTTCGTTGTG 65
 |||||

QY 61 CTGAGAACACGTAAGCACTAAGCTGTGTGAGAGAAA---CTTGTCTTAAAGAACTGCAGC 117
 |||||
 Db 66 CTGAGAACACCTAAGCACTAAGCTGTGTGAGAGAACTCTTGTGTAAGAAACCGCAGC 125
 |||||

QY 118 TTCTGGGCTCAGGGGTGCAGACCCCTCCGCTAGCTCCAGACCGCTGTGACACGACGAC 177
 |||||
 Db 126 TTCTGGAGCCAGGGGTGCAGACCCCTCCAGAGTTCACGACCGCTGTGACACGACGAC 185
 |||||

QY 178 TCCTTAATGACACGCTGCCATGTAAGCGACCTGTAACCTTATCAGCCCATGCTCAATACGT 237
 |||||
 Db 186 TCCTTAATGACACGCTGCCATGTAAGCGACCTGTAACCTTATCAGCCCATGCTCAATACGT 245
 |||||

QY 238 AACTTTGTACTGTACGTAC---GATGGGTGTAAAGAGCTCTGCTCTTGTATTCATAGTG 294
 |||||
 Db 246 AACTTTGTACTGTACGTACGATGATGATGTGTGACAGCTCAGCTCTTGTATTCATAGTG 305
 |||||

QY 295 AGTTCTCTAAAATACCAGCTGACCGGCTTCTGAGGCTTTGACAGAACTCTGGCTCCTG 354
 |||||

Db 306 AGTTTCTAAAAGACGACGCGACTAGCTTTTGCAGACTTTGAACAGAACTCTGGTTCCTG 365
 |||||

QY 355 TGTTCCTCTAACGAAGATTCTGTTCCTAGTCGTTGGGTGTCTGGTGGAGTGTGTGAA 414
 |||||

Db 366 TGTTCCTCTAACGAAGATTCTGTTCCTAGTCGTTGGGTGTCTGGTGGAGTGTGTGAA 423
 |||||

QY 415 ACACGACGCTCATCAAGGAGAGACAGACAGTATTTTCTACTAATATGAAGTAGAGTAATTT 474
 |||||

Db 424 ACACGACATGATCAAGGAGAGACAGACAGTATTTTCTACTAATATGAAGTAGAGTACITT 483
 |||||

QY 475 ACACCTAGATTGTACATGAGGTAATTTCACTGAATAAAAGTGTACGGGTAAAGCTTTT 534
 |||||

Db 484 ACACCTAGATTGTACATGAGGTAATTTCACTGAATAAAAGTGTACGGGTAAAGCTTTT 543
 |||||

QY 535 ACGGTTAATTTCTGCAAAACAGTAGATGACAAATGGCCGATCTTATCAGTGTCTC----- 589
 |||||

Db 544 ACGGTTAATTTCTGCAAAACAGTAGTGGTGAATAATGGCCGATCTTATCAGTGTCTCTTGAG 603
 |||||

QY 590 -----TCTTGAGCCGCCCTTCCCTCTGCTGTCTCCCTCCAGATGGGGCGTTGAGTCCA 642
 |||||

Db 604 CATCCCTTGGCGCCCCCACCCTCCACTCCGCTCCTCCAGATGGGGCGTTGAGTCCA 663
 |||||

QY 643 TATTAACTGGCCATCTCTACAGTTGCTAACTTAGCAAGTGTCTTTTAGGACCCCT 702
 |||||

Db 664 TATTAACTGGCCATCTCTACAGTTGCTAACTTAGCAAGTGTCTTTTAGGACCCCT 720
 |||||

QY 703 TTCTTAACGAGCAATATGTCTGACCTGTACT 733
 |||||

Db 721 TTCTTAACGAGCAATATGTCTGACCTGTACT 751
 |||||

RESULT 8
 BUS11766
 LOCUS
 DEFINITION
 AGENCOURT 10114251 NIH_MGC_134 Mus musculus cDNA clone
 IMAGE:6506583 5', mRNA sequence.
 BUS11766
 ACCSSION
 VERSION
 KEYWORDS
 EST.
 BUS11766.1 GI:22817999
 SOURCE
 Mus musculus (house mouse)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 960)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Dr. David Rowe
 cDNA Library preparation: Invitrogen Corp
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM14069 row: e column: 16
 High quality sequence stop: 669.
 Location/Qualifiers
 1. 960
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:6506583"
 /tissue_type="undifferentiated limb"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC 134"
 /note="Vector: pCMV-SPORT6.1.cdb; Site 1: EcoRV; Site 2:
 NotI. Cloned unidirectionally. Primer: Oligo dT. Average
 insert size 1.7 kb. Constructed by ResGen, Invitrogen
 Corp. Note: this is a NIH MGC Library."
 246 a 218 c 196 g 298 t 2 others

BASE COUNT 246 a 218 c 196 g 298 t 2 others

FEATURES
 source

ORIGIN	Query Match	54.3%;	Score 552.2;	DB 13;	Length 960;
KEYWORDS	Best Local Similarity	86.6%;	Pred. No. 2.3e-147;		
SOURCE	Matches 733;	Conservative 0;	Mismatches 79;	Indels 34;	Gaps 10;
ORGANISM					
QY	45	ACTAAGTTCGGTGTGGCTGAGAACAGCTAAGCACTAAGCTGTGTGAGAC---	TTTGTCC	101	
Db	3	AATAAGTTCGGTGTGGCTGAGAACACCTAAGCACTAATACTCTCTTGAGAGACTTCTGTG	62		
QY	102	TTAAGACACTGACGCTCTGGGCTCAGGGGTGACAGCCCTCCCGTAGCTCCAGACCGT	161		
Db	63	CTAAGAACCGCAGCTTCTGAGCCGAGGGGTGACAGCCCTCCAGAGTTCACAGACCGT	122		
QY	162	GTGACACAGCAGCCTCTTAATACACAGCTGCCATGTAAGCAGCCTGTAACTTATCAG	221		
Db	123	GTGACACAGCAGCCTCTTAATACACAGCTGCCATGTAAGCAGCCTGTAACTTATCAG	182		
QY	222	CCCATGCTCATTAACCTTGTACTGTACGTACCTAC---GATGGGTGTAAACAGCTCTGCT	278		
Db	183	CCCATGCTCATTAACCTTGTACTGTACGTACGTACGTACGTACGTACGTACGTACGT	242		
QY	279	CTTTGATTTATAGTGTCTCTAAATACCAGCTGACCGCTCTTCAGAGCTTTGAAC	338		
Db	243	CTTTGATTTATAGTGTCTCTAAATACCAGCTGACCGCTCTTCAGAGCTTTGAAC	302		
QY	339	AGAACTCTGGCTCTGTGGTCTCTAAGCAAGTATCTGCTTCTCTAGTCTGGTGTGCT	398		
Db	303	AGAACTCTGGTCTGTGGTCTCTAAGCAAGTATCTGCTTCTCTAGTCTGGTGTGCT	362		
QY	399	GGGTGGAGTGTGAACACAGCTCATCAAGAGGACAGACAGATTTTGTACTAATATG	458		
Db	363	GGGTGGA--GTGTGAACACAGCTCATCAAGAGGACAGACAGATTTTGTACTAATATG	420		
QY	459	AAGTAGAGATTAATTTACACTACATGTACATGGAGTAATTCACCTGAATAAAGTGTCA	518		
Db	421	AAGTAGAGATTAATTTACACTACATGTACATGGAGTAATTCACCTGAATAAAGTGTCA	480		
QY	519	CGGTAAGAGCTTTTAAACGGTTAATTTCTGTCAACAGTAGATGACAAATGGCGATCTT	578		
Db	481	CGGTAAGAGCTTTTAAACGGTTAATTTCTGTCAACAGTAGATGACAAATGGCGATCTT	540		
QY	579	ATCAGTGTCTC-----TCTTGAGCCCTCTTCCCTCTGCTGTCTCCCTCCACAG	626		
Db	541	ATCAGTGTCTCTTGAGATCCCTTCGCGCCCTCCACCTCCACTGCGCTCCCTCCACAG	600		
QY	627	TGGGGGCTTGAGTCCATATTTAACTGGCCATCTCCACAGTTCCTAATAGCAAGTGTCT	686		
Db	601	TGGGGGCTTGAGTCCATATTTAACTGGCCATCTCCACAGTTCCTAATAGCAAGTGTCT	659		
QY	687	TTTCTTTAGACCCCTCTTACAGGACATATGTCTGACCTGTACTATTAAGATCTTTCT	746		
Db	660	TTTCTTTA--GAACCCCTCTTACAGGACATATGTCTGACCTGTACTATTAAGATCTTTCT	718		
QY	747	GATAATGCA-----TTCGGAGATTTTGTGTAGATAGTAGAAGTGCCTCTCTG--TTT	798		
Db	719	GATAACCGACAGATTTTGTGTGTAGATAGTAGAAGTGCCTCTCTG--TTT	778		
QY	799	TCACCTTCTTTTACTC---AGCTGACATAGCTTCCCTTCTGTTT--TCTAGTACCTGGTG	854		
Db	779	TCGCTTACTTTTACTCAGAAGCTAGTTAGTGTCTTCTCTAGTTCCTTCTAGTTCCTAG	838		
QY	855	TAGAAA 860			
Db	839	TGCAAA 844			
RESULT 9	BG665146	557 bp	mRNA	linear	EST 30-APR-2001
LOCUS	DRABZ12	Rat DRG Library	Rattus norvegicus	cdna clone	DRABZ12 5',
DEFINITION					mRNA sequence.
ACCESSION	BG665146				

BG665146.1	GI:13887068
EST.	
KEYWORDS	Rattus norvegicus (Norway rat)
SOURCE	Rattus norvegicus
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
REFERENCE	
AUTHORS	1. (bases 1 to 557) Xiao H.S., Huang Q.H., Zhang F.X., Bao L., Lu Y.J., Guo C., Yang L., 'Huang, W.J., Fu G., Xu S.H., Cheng X.P., Yan Q., Zhu Z.D., Zhang X., Chen Z., Han Z.G. and Zhang X. Identification of gene expression profile of dorsal root ganglion in the rat peripheral axotomy model of neuropathic pain Proc. Natl. Acad. Sci. U.S.A. 99 (12), 8360-8366 (2002)
TITLE	22056133
JOURNAL	12060780
MEDLINE	Contact: Zhang Xu
PUBMED	Laboratory of Sensory System
COMMENT	Institute of Neuroscience 320 Yue Yang Road, Shanghai 200031, P.R.China Tel: 86-21-64748700-121 Fax: 86-21-64713446 Email: xu.zhang@ion.ac.cn This clone is also available at Chinese National Human Genome Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong New Area, P.R.China. Please contact with Zhang Xu (xu.zhang@ion.ac.cn) or Han Zeguan (hanzg@chgc.sh.cn)
FEATURES	PCR Primers FORWARD: T3 BACKWARD: T7 Seq primer: T3 POLYA=No.
Location/Qualifiers	1..557 /organism="Rattus norvegicus" /mol_type="mRNA" /strain="Sprague-Dawley" /db_xref="taxon:10116" /clone="DRABZ12" /sex="male" /tissue_type="dorsal root ganglion" /dev_stage="adult" /clone_lib="Rat DRG Library" 142 a 130 c 128 g 156 t 1 others
BASE COUNT	142 a 130 c 128 g 156 t 1 others
ORIGIN	
Query Match	54.2%; Score 551.2; DB 10; Length 557;
Best Local Similarity	99.3%; Pred. No. 3.2e-147;
Matches 553;	Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY	55 GTTGTGCTGAGAACAGCTAAGCACTAAGCTGTGTGAGAGACTTTGTCTTAAAGAGACTGC 114
Db	1 GTTGTGCTGAGAACAGCTAAGCACTAAGCTGTGTGAGAGACTTTGTCTTAAAGAGACTGC 60
QY	115 AGTTCTTGGGCTCAGGGGTGACAGCCCTCCGCTAGCTCCAGACCGGTGACACAGCACA 174
Db	61 AGTTCTTGGGCTCAGGGGTGACAGCCCTCCGCTAGCTCCAGACCGGTGACACAGCACA 120
QY	175 GCCTCTCTTAATGACAGCTGCCATGTAAACGCACTGTAACTATCAACCCATGTCTATTA 234
Db	121 GCCTCTCTTAATGACAGCTGCCATGTAAACGCACTGTAACTATCAACCCATGTCTATTA 180
QY	235 CGTAACTTTGTACTGTACGTACAGTGGGTGTAAACAGCTCTGCTCTTTGATTTTCAATAGT 294
Db	181 CGTAACTTTGTACTGTACGTACAGTGGGTGTAAACAGCTCTGCTCTTTGATTTTCAATAGT 240
QY	295 AGTTCTCTAAATACAGCTGACCGCTCTCGAGGCTTTGAACAGAACTCTTGCTCTCTG 354
Db	241 AGTTCTCTAAATACAGCTGACCGCTCTCGAGGCTTTGAACAGAACTCTTGCTCTCTG 300
QY	355 TGTTCCTCTTAAACGAACTATTCTGCTTCTTCTAGTCTGGGTGTGCTGGGTGGAGTGTGAAA 414
Db	301 TGTTCCTCTTAAACGAACTATTCTGCTTCTTCTAGTCTGGGTGTGCTGGGTGGAGTGTGAAA 360

415 ACACGACGTCATCAAGAGAGACAGACAGATTTTACCTAATATGAAGTAGAGATTAAATTT 474
 Db |||||||
 361 ACACGACGTCATCAAGAGAGACAGACAGATTTTACCTAATATGAAGTAGAGATTAAATTT 420
 QY |||||||
 475 ACACCTACATTTGACATGAGTAATCAACTGAATAAAAAGTGTCAAGGGTAAAGCTTTTAA 534
 Db |||||||
 421 ACACCTACATTTGACATGAGTAATCAACTGAATAAAAAGTGTCAAGGGTAAAGCTTTTAA 480
 QY |||||||
 535 ACGGTTAATTTCTGTCAACAGTAGATGACAAATGGCCGATCTTATCAAGTGTCTCTCTTTG 594
 Db |||||||
 481 ACGGTTAATTTCTGTCAACAGTAGATGACAAATGGCCGATCTTATCAAGTGTCTCTCTTTG 540
 QY |||||||
 595 AGCCGCCCTTCCCTCG 611
 Db |||||||
 541 AGCCGCCCTTCCCTCG 557

CA319273 793 bp mRNA linear EST 26-NOV-2002
 UI-M-FW0-cbz-a-24-0-UI-r1 NIH_BMAP_FW0 Mus musculus cDNA clone
 IMAGE: 6816385 5', mRNA sequence.
 CA319273
 CA319273.1 GI:24537397
 EST.
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 793)
 NIH-MGC <http://mgs.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)

Seq primer: pYX-5.
 Location/Qualifiers
 1. 793
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE: 6816385"
 /tissue_type="whole brain"
 /dev_stage="embryo 13.5, 14.5, 15.5, 16.5, 17.5 dpc"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_lib="NIH_BMAP_FW0"
 /note="Organ: Brain; Vector: pYX-Asc; Site: 1: EcoR I;
 Site 2: Not I; The library was constructed according
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
 1996. Denatured RNA was size fractionated on a 1% agarose
 gel. First strand cDNA synthesis was primed with oligo-dT
 primer containing a Not I site. Double strand cDNA was
 size selected according to mRNA size fraction, ligated
 with EcoR I adaptor, digested with NotI and then cloned
 directionally into pYX-Asc vector. The library tag
 sequence located between the Not I site and the polyA tail
 is AGCAGACAG. This library was created for the University
 Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
 Developing Mouse Nervous System', supported by National
 Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
 program coordinator."

200 a 185 c 165 g 242 t 1 others

BASE COUNT

ORIGIN

Query Match 52.8%; Score 537.4; DB 14; Length 793;
 Best Local Similarity 87.2%; Pred. No. 3.7e-143;
 Matches 677; Conservative 0; Mismatches 72; Indels 27; Gaps 7;

QY 64 AGAACACGTAAGCACTAAGCTGTGTGAGAGA---CTTTGTCTTAAAGAGACGTGAGCTTC 120
 Db 1 AGAACACGTAAGCACTAAGCTGTGTGAGAGA---CTTTGTCTTAAAGAGACGTGAGCTTC 60
 QY 121 TGGGCTCAGGGGTGCAGACCCCTCCGCTAGCTCCAGACCGGTGTGACAGACGACGCTCC 180
 Db 61 TGGAGCCAGGGGTGCAGACCCCTCCAGAGTCCAGACCGGTGTGACAGACGACGCTCC 120
 QY 181 TTAATGACACGCTGCCATGTAAACGACCTGTAACTTATCAGCCCATGCTCATTAAGTAAC 240
 Db 121 TTAATGACACGCTGCCATGTAAACGACCTGTAACTTATCAGCCCATGCTCATTAAGTAAC 180
 QY 241 TTTGTACTGTAGCTCAC---GATGGGTGTAAACGCTCTGTCTTTGATTTTATAGTAGT 297
 Db 181 TTTGTACTGTAGCTCACAGTGTAGTGTGACGCTCAGCTCTTTGATTTTATAGTAGT 240
 QY 298 TCTCTAAATACACGCTGACCGGCTTCTGAGGCTTTTGAACAGAACTCTGCTCTCTGTGT 357
 Db 241 TTTCTAAAGACGACCGGCTAGCTTTTGCAGACTTTTGAACAGAACTCTGCTCTCTGTGT 300
 QY 358 TGCCTCTAAACGAAGTATTCTCTTCTAGTCTGGGTGTGCTGGGTGGAGTGTGTGAACA 417
 Db 301 TGCCTCTAAACGAAGTATTCTCTTCTAGTCTGGGTGTGCTGGGTGGAGTGTGTGAACA 358
 QY 418 CGAGTCTAATCAAGAGACAGACAGATTTTGTAGCTAATATGAAGTAGAGATTAATTACA 477
 Db 359 CGAGTCTAATCAAGAGACAGACAGATTTTGTAGCTAATATGAAGTAGAGATTAATTACA 418
 QY 478 CTACATTTGATCATGAGTAATCAACTGAATAAAGTGTCAAGGTGAAGCTTTTAAAG 537
 Db 419 CTACATTTGATCATGAGTAATCAACTGAATAAAGTGTCAAGGTGAAGCTTTTAAAG 478
 QY 538 GTTAATTTCTTCAACAGTAGATGACAAATGGCGGATCTTTATCAGTGTCTC----- 589
 Db 479 GTTAATTTCTTCAACAGTAGATGACAAATGGCGGATCTTTATCAGTGTCTCCTTGAGCAT 538
 QY 590 ----TCTTGAGCCGCCCTTCCCTCTGCTGCTCCCTCCAGATGGGCGGTGAGTCCATAT 645
 Db 539 CCCTTCGCGCCGCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTAT 598
 QY 646 TTAACCTGGCCATCTCAGCTGTGCTAACTTAGCAAGTCTTTCTTTAGGACCCCTTC 705
 Db 599 TTAACCTGGCCATCTCAGCTGTGCTAACTTAGCAAGTCTTTCTTTA-GAACCCCTTC 656
 QY 706 TTAACGAGCAATATGTCTGACCTGTACTATAAGATCTTTCTGATAATGCA-----TTCGG 760
 Db 657 TTAACGAGCAATATGTCTGACCTGTACTATAAGATCTTTCTGATAATGCAAGATTTT 716
 QY 761 AGATTTTGTGTAGATAGTAGAGTGGTCTCTGTTTCCCTCCCTCCCTCCCTTACTCAG 816
 Db 717 GTTTTGTGTAGTCTCAGTANAAGTGGCTCTCTGTTTCCCTTACTTACTCAG 772

RESULT 11
 LOCUS BI646446
 DEFINITION 603279026f1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5319127 5',
 mRNA sequence.
 ACCESSION BI646446
 VERSION BI646446.1 GI:15560682
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 897)
 NIH-MGC <http://mgs.nci.nih.gov/>.

TITLE
JOURNAL
COMMENT

National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

Plate: LLAM1803 row: d column: 08
High quality sequence stop: 890.

FEATURES

source

1. .897
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="129,C57BL/6J,FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5319127"
/tissue_type="tumor, gross tissue"
/dev_stage="10 months"
/lab_host="DH10B"
/clone_lib="NCI_CGAP Mam3"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site.1: Sali;
Site.2: NotI; Cloned unidirectionally. Primer: Oligo dr.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
Reference for transgenic model: Xu et al., Nature Genetics
22, 37-43 (1999)."

BASE COUNT 228 a 213 c 190 g 266 t
ORIGIN

Query Match 52.6%; Score 535.2; DB 12; Length 897;
Best Local Similarity 83.6%; Pred. No. 1.7e-142;
Matches 734; Conservative 0; Mismatches 113; Indels 31; Gaps 10;

QY 60 GCTGAGAACACGTAAAGCACTAACTCTCTTGAGAGACTTCTGTTGCTTAAGAACGCGCAG 116
DB 1 GCTGAGAACACGTAAAGCACTAACTCTCTTGAGAGACTTCTGTTGCTTAAGAACGCGCAG 60

QY 117 CTCTGGCTCAGGGGTGAGACGCTCCCTAGCTCCAGACCGTGTGACACAGCACAGC 176
DB 61 CTCTGGAGCCA-GGGTGCAGGCGCTTCCAGAGTTCACAGACCGTGTGACACAGCACAGC 119

QY 177 CTCCTTAATGACACGCTGCCATGTAAAGCACTGTAACTTATCAGCCCATGCTCATTAACG 236
DB 120 CTCCTTAATGACACGCTGCCATGTAAAGCACTGTAACTTATCAGCCCATGCTCATTAACG 179

QY 237 TAACCTTTGTAAGTACGTAC---GATGGGTGTAACAGCTCTGCTCTTTGATTTCATAGT 293
DB 180 TAACCTTTGTAAGTACGTACGTACAGTGATGAGTGTGACAGCTCAGCTCTTTGATTTCATAGT 239

QY 294 GAGTTCTTAATAATACAGTGCAGCGCTTCTCCAGGCTTTGACAGAACTCTGGCTCCT 353
DB 240 GAGTTTCTAAAAGACAGCGAGTACTGTTTCAGACTTTTGAACAGAACTCTGGTTCCT 299

QY 354 GTGTGGCTCTAACGAAGTATTCGTCTCCAGTGTGGGTGTGCTGGGTGGAGTGTGA 413
DB 300 GTGTGGCTCTAATGAATATTCGTCTCCAGTGTGGGTGTGCTGGGTGGAGTGTG--A 357

QY 414 AACACGAGTCAATCAAGGAGACAGACAGTATTTTGACTAATATGAAGTAGAGATTAAAT 473
DB 358 AACACGAGTCAATCAAGGAGACAGACAGTATTTTGACTAATATGAAGTAGAGATTAAAT 417

QY 474 TACACTACATTTGACATGAGTAACTCACTGTAATAAAGTGTACAGGGTAAAGCTTTT 533
DB 418 TACACTACATTTGACATGAGTAACTCACTGTAATAAAGTGTACAGGGTAAAGCTTTT 477

QY 534 AACGGTTAAATTTCTGTCAACAGTAGATGACAAATGGCCGATCTTATCAGTGTCTC--- 589
DB 478 AACGGTTAAATTTCTGTCAACAGTAGATGACAAATGGCCGATCTTATCAGTGTCTTGA 537

QY 590 -----TCTTGAGCCCCCTTTCCCTCTGTCTCCCTCCAGATGGGGGTGAGTCC 641
DB 538 GCATCCCTTCGGGCCCCCCCCACCTCCACTCCGCTCCCTCCAGATGGGGGTGAGTCC 597

QY 642 ATATTAAACT-GGCACTCTCAGTGTCTAACTTAGCAAGTGCCTTTCTTTAGGACCC 700
DB 598 ATATTAAACTGGGCCACCCCTCAGTGTCTAACTTAGCAAGTGCCTTTCTTTA-GAACC 656

QY 701 CTTTCTTAACGACAATATGTCTGACTGTACTATTAAGATCTTTCTGATAATGCA---T 756
DB 657 CTTTCTTAACGACAATATGTCTGACTGTACTATTAAGATCTTTCTGATAACGACAGA 716

QY 757 TCGGAGATTTTTTGGTAGATAGTAGAGTGGTCTCTGTTTTCACCTTCTTTACTACAG 816
DB 717 TTTGGTCTCTCTGGTAGCTCAGTAGAAGTGGCTCTGTTTTCGCTTACTTTACTCAG 776

QY 817 -CTGACTAGTGTCTCCCTTCGTTTCTAGTAACTGGGTGT---AGAAATCAGTGTCTGG 872
DB 777 AGCTAGTGTGTCTCCCTTAGTTCCTAGCCAACTAGAGTGTGCCAAATCATGTGTTGCG 836

QY 873 GCTTTACAGTTTTTAAACTATTATTAGATATTCTGAAAC 910
DB 837 CGCCTTACGGGTTAAACTATTATTAGATATTCTTAAC 874

RESULT 12
BI851489
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 714)
NIH-MGC http://img.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

Plate: LLAM1994 row: e column: 24
High quality sequence stop: 714.
Location/Qualifiers
1. .714
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N-3"
/db_xref="taxon:10090"
/clone="IMAGE:5390207"
/tissue_type="tumor, biopsy sample"
/dev_stage="5 months"
/lab_host="DH10B"
/clone_lib="NCI_CGAP Mam2"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site.1: Sali;
Site.2: NotI; Cloned unidirectionally. Primer: Oligo dr.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

FEATURES

source

BASE COUNT 187 a 172 c 149 g 206 t
ORIGIN

Query Match 52.0%; Score 528.6; DB 12; Length 714;
Best Local Similarity 89.5%; Pred. No. 1.2e-140;
Matches 642; Conservative 0; Mismatches 54; Indels 21; Gaps 6;

QY 2 CCTATTCTTCTGCTCAGATTAGATTGCGAAATACCTTGTAAGTTCAGTTCGTTGTC 61
 Db 1 CCCATTCTTCTGCTCAGATTAGATTGCGAAATACCTTGTAAGTTCAGTTCGTTGTC 60
 QY 62 TGAGAACACCTGTAAGCATTAGCTCTTGAGAGA---CTTTGTCCTTAAGAGAGCTGACGT 118
 Db 61 TGAGAACACCTGTAAGCATTAGCTCTTGAGAGACTTCTGTTGCTAAGAGAGCGCAGCT 120
 QY 119 TCTGGGCTCAGGGGTGAGACCCCTCCCTAGCTCCGAGACCGGTGTGACAGACAGCCCT 178
 Db 121 TCTGGAGCCA-GGGTGCAGACCCCTCCGAGAGTCCGAGACCGGTGTGACAGACAGCCCT 179
 QY 179 CCTTAATGACAGCTGCGATGTAACGACCTGTAACCTTATCAGCCCATGCTCAATAGTA 238
 Db 180 CCTTAATGACAGCTGCGATGTAACGACCTGTAACCTTATCAGCCCATGCTCAATAGTA 239
 QY 239 ACTTTGACTGTAGCTCAC---GATGGGTGTAACAGCTCTGCTCTTTGATTTCATAGTA 295
 Db 240 ACTTTGACTGTAGCTCACAGTGTAGTGTGACAGCTCAGCTCTTTGATTTCATAGTA 299
 QY 296 GTTCTCTAAATACCAGCTGACCGGCTTCTGACGGCTTGAACAGAACTCTGGCTCCTGT 355
 Db 300 GTTTTCTAAAGACGACCGGCTAGCTTTTGCAGACTTGAACAGAACTCTGGTTCCTGT 359
 QY 356 GTTTCCTCTAACGAAGTATTCTGTTCTAGTCTGCGGGTGTGCTGGGTGAGTGTGAAA 415
 Db 360 GTTTCCTCTTAATGAAGTATTCTGTTCTAGTGTGCGGGTGTGCTGGGTGGA--GTTGAAA 417
 QY 416 CAGACGCTCATCAAGGAGACAGACAGTATTTTCACTAATATGAAGTAGAGATTAAATTA 475
 Db 418 CAGACATGATCAAGAGAGACAGACAGTATTTTGACTAATATGAAGTAGAGATTACTTTA 477
 QY 476 CACTACATTTGACATGAGTAATTCACATGAATAAAGTGTACCGGTAAGCTTTTAA 535
 Db 478 CACTACATTTGACATGAGTAATTCACATGAATAAAGTGTACCGGTAAGCTTTTAA 537
 QY 536 CGGTAAATTTCTGCAACAGTAGATACAAATGCGCGATCTTATCAGTGCT----- 588
 Db 538 CGGTAAATTTCTGCAACAGTAGATGACAAATGCGCGATCTTATCAGTGCTTTGAGC 597
 QY 589 ----CTTCTGAGCCCTCTCCCTCTGTCCTCCCTCCGATGGGCGTTGAGTCCATA 644
 Db 598 ATCCCTTGGGGCCCCCACCCTCCTGCTGCTCCCTCCGATGGGCGTTGAGTCCATA 657
 QY 645 TTTAACTGGCCATCTTCAAGTTGCTAACTTAGCAAGTGCTTTCTTTAGGACCCC 701
 Db 658 TTTAACTGGCCATCTTCAAGTTTCTAACTTAGCAAGTGCTTTCTTTAGAACCCC 713

RESULT 13
 BQ931374
 LOCUS
 DEFINITION BQ931374 915 bp mRNA linear EST 21-AUG-2002
 AGENCOURT 8950013 NCI CGAP Mam2 Mus musculus cDNA clone
 IMAGE:6442308 5', mRNA sequence.
 ACCESSION BQ931374
 VERSION BQ931374.1
 KEYWORDS GI:22346405
 SOURCE EST.
 ORGANISM Mus musculus (house mouse)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus.
 1 (bases 1 to 915)
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LIA31966 row: o column: 13
 High quality sequence stop: 669.

FEATURES

source
 1..915
 Location/Qualifiers
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N-3"
 /db_xref="taxon:10090"
 /clone="IMAGE:6442308"
 /tissue_type="tumor, biopsy sample"
 /dev_stage="5 months"
 /lab_host="DH10B"
 /clone_lib="NCI CGAP Mam2"
 /notes="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
 Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigator
 providing samples: Gilbert Smith, NIH"
 BASE COUNT 231 a 207 c 193 g 283 t 1 others
 ORIGIN

Query Match 52.0%; Score 528.6; DB 13; Length 915;
 Best Local Similarity 84.8%; Pred. No. 1.4e-140;
 Matches 696; Conservative 0; Mismatches 94; Indels 31; Gaps 8;
 QY 116 GTTCTGGGCTCAGGGGTGAGACCCCTCCGTTAGCTCCGAGACCGGTGTGACAGACAG 175
 Db 10 GGTTCCTGAGCCAGGGGTGACAGCCCTTCCAGAGTTCGAGACCGGTGTGACAGACAG 69
 QY 176 CCTCCTTAATGACAGCGCTGCATGTACCGCACCTGTAACTTATCAGCCCATGCTCATTA 235
 Db 70 CCTCCTTAATGACAGCGTGGCCATGTAAAGCACCTGTAACTTATCAGCCCATGCTCATTA 129
 QY 236 GTAACCTTGTACTGTACGTAC---GATGGGTGTAACAGCTCTGCTCTTTGATTTCATAG 292
 Db 130 GTAACCTTGTACTGTACGTACAGTGTAGTGTGACAGCTCAGCTCTTTGATTTCATAG 189
 QY 293 TGAGTCTCTAAATAACAGCTGACCGGCTTCTGACGGCTTTGAACAGAACTCTGGGTCC 352
 Db 190 TGAGTCTCTAAAGACACCGGCTAGCTTTTGCAGACTTTGAACAGAACTCTGGGTCC 249
 QY 353 TGTGTTGCCCTTAACGAAGTATTCTGTTCTAGTCTGTTGGGTGCTGGTGGAGTGTGTG 412
 Db 250 TGTGTTGCCCTTAATGAAGTATTCTGTTCTAGTGTGTTGGGTGCTGGTGGAGTGTGTG 307
 QY 413 AAACACGACGCTCATCAAGGAGACAGACAGTATTTTCACTAATATGAAGTAGAGATTAA 472
 Db 308 AAACACGACATGATCAAGGAGACAGACAGTATTTTCACTAATATGAAGTAGAGATTACT 367
 QY 473 TTACACTACATTTGTACATGAGTAATTCACATGAATAAAGTGTACGGGTAAAGCTTTT 532
 Db 368 TTACACTACATTTGTACATGAGTAATTCACATGAATAAAGTGTACGGGTAAAGCTTTT 427
 QY 533 TAACGGTTAATTTCTGTCAACAGTAGATGACAAATGCGCGATCTTATCAGTGCTC--- 589
 Db 428 TAACGGTTAATTTCTGTCAACAGTAGATGACAAATGCGCGATCTTATCAGTGCTCTTTG 487
 QY 590 -----TCTTGAGCCCTCTCCCTCTGCTGCTCCCTCCAGATGGGCGTGTAGTC 640
 Db 488 AGCATCCCTTGGCGCCCCCACCCTCAGCTGCGCTCCCTCCAGATGGGCGTGTAGTC 547
 QY 641 CATATTAAACTGGCCATCTCAGTGTCTAACTTAGCAAGTGTCTTTCTTTAGGACCC 700
 Db 548 CATATTAAACTGGCCATCTCAGTGTCTAACTTAGCAAGTGTCTTTCTTTA-GAACC 605
 QY 701 CCTTCTTAACGAGCAATATGCTGACCTGTACTATAGATCTTTCTGATATGA----- 755
 Db 606 CCTTCTTAACGAGCAATATGCTGACCTGTACTATAGATCTTTCTGATATGAACGAGGA 665
 QY 756 --TTCGAGATTTTTTTTGGTGTAGATAGTAGAGTGGCTCTCTGTTTTCACCTTCTCTTACT 813
 Db 666 TTTTGTGTTTTTTTTTGGTAGCTCAGTAGAGTGGCTCTCTGTTTTCGCTTCTTACT 725

FEATURES
Source

Location/Qualifiers

1. 639
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-DQ1-clt-n-08-0-UI"
/dev_stage="adult"
/lab_host="DHI0B (Life Technologies)"
/clone_lib="UI-R-DQ1"
/note="Vector: pUT3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-DQ1
library is a normalized Rat Cell line R3327-5A library
(nr5A) constructed in p3T7 PAC vector according to the
procedure described by Bonaldo, Lennon & Soares
(Normalization and Subtraction: Two Approaches to
Facilitate Gene Discovery. Genome Research 6: 791-806,
1996). The oligonucleotide used to prime first strand
synthesis contained the sequence tag GGAAGTATC between
the Not I cloning site and dT18 stretch. The Rat Cell line
R3327-5A was provided by Mary Hendrix of the University of
Iowa.

TAG_LIB=UI-R-DQ1
TAG_TISSUE=rat cell line R3327-5a
TAG_SEQ=GGACTAGATC"
BASE COUNT 183 a 130 c 136 g 190 t
ORIGIN

Query Match 51.7%; Score 526; DB 13; Length 639;
Best Local Similarity 100.0%; Pred. No. 6.3e-140;
Matches 526; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCCTATCTGCTCAGATTAGATTCGCCAAATACCTTGTGAACCTAAGTTGGTTGTG 60
Db 541 CCCCTATCTGCTCAGATTAGATTCGCCAAATACCTTGTGAACCTAAGTTGGTTGTG 482

QY 61 CTGAGAACAGCTAGACTAAGCTGTTGAGAGACTTTGTCTTAAAGAGACTGCAGCTTC 120
Db 481 CTGAGAACAGCTAGACTAAGCTGTTGAGAGACTTTGTCTTAAAGAGACTGCAGCTTC 422

QY 121 TGGCTCAGGGGTGCAGACCCCTCCGCTAGCTCCAGACCGGTGTCACACAGACGCTCC 180
Db 421 TGGCTCAGGGGTGCAGACCCCTCCGCTAGCTCCAGACCGGTGTCACACAGACGCTCC 362

QY 181 TTAATGACAGCTGCCATGTAAAGCAGCTGTAACTTATCAGCCCATGCTCATTTACGTAAC 240
Db 361 TTAATGACAGCTGCCATGTAAAGCAGCTGTAACTTATCAGCCCATGCTCATTTACGTAAC 302

QY 241 TTTGTACTGTACGTACGATGGGTGTAAACAGCTCTGCTCTTTGATTTTATAGTGAGTTCT 300
Db 301 TTTGTACTGTACGTACGATGGGTGTAAACAGCTCTGCTCTTTGATTTTATAGTGAGTTCT 242

QY 301 CTAAATACAGCTGACCGGCTTCTGCAGGCTTTGAAACAGAACTCTGGCTCCTGTGTTC 360
Db 241 CTAAATACAGCTGACCGGCTTCTGCAGGCTTTGAAACAGAACTCTGGCTCCTGTGTTC 182

QY 361 CTCTACGAGTATCTGTTCTTCTAGTCTGGGTGTGCTGGGTGAGTGTGTGAAACAGA 420
Db 181 CTCTACGAGTATCTGTTCTTCTAGTCTGGGTGTGCTGGGTGAGTGTGTGAAACAGA 122

QY 421 CGTCATCAAGGAGACAGACAGTATTTTGAATATATGAAGTAGAGATTAAATTACACTA 480
Db 121 CGTCATCAAGGAGACAGACAGTATTTTGAATATATGAAGTAGAGATTAAATTACACTA 62

QY 481 CATGTACATGGAGTAAATCAACTGAATAAAAGTGTACCGGTAAA 526
Db 61 CATGTACATGGAGTAAATCAACTGAATAAAAGTGTACCGGTAAA 16

Search completed: November 23, 2003, 15:50:06
Job time : 2395.73 secs

This Page Blank (uspto)

GenCore version 5.1.6
 Copyright (c) 1993 - 2003 CompuGen Ltd.
 OM nucleic - nucleic search, using sw model
 Run on: November 23, 2003, 10:50:08 ; Search time 77.549 Seconds
 (without alignments)
 5788.425 Million cell updates/sec
 Title: US-09-717-321A-15
 Perfect score: 1017
 Sequence: 1 cccattcttctcagatt.....cctttgggtctgagggttc 1017
 Scoring table: IDENTITY NUC
 Gapop 10.0 , Gapext 1.0
 Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : Issued Patents NA: *
 1: /cgn2_6/prodata/2/ina/5A_COMB.seq.*
 2: /cgn2_6/prodata/2/ina/5B_COMB.seq.*
 3: /cgn2_6/prodata/2/ina/6A_COMB.seq.*
 4: /cgn2_6/prodata/2/ina/6B_COMB.seq.*
 5: /cgn2_6/prodata/2/ina/6C_COMB.seq.*
 6: /cgn2_6/prodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	187.2	18.4	605	3	US-09-385-982-483
C 2	68	6.7	588	3	US-09-385-982-133
C 3	54.2	5.3	7218	1	US-08-232-463-14
C 4	35.4	3.5	2038	1	US-08-181-271A-1
C 5	35.4	3.5	2038	1	US-08-449-515-1
C 6	35.4	3.5	2038	1	US-08-444-803-1
C 7	35.4	3.5	2038	1	US-08-449-043-1
C 8	35.4	3.5	2038	1	US-08-456-265A-1
C 9	35.4	3.5	2038	1	US-08-455-416-1
C 10	35.4	3.5	2038	1	US-08-455-244-1
C 11	35.4	3.5	2038	1	US-08-454-876-1
C 12	35.4	3.5	2038	2	US-08-457-364-1
C 13	35.4	3.5	2038	2	US-08-456-262-1
C 14	35.4	3.5	2038	2	US-08-456-240-1
C 15	35.4	3.5	2038	2	US-08-455-736-1
C 16	35.4	3.5	2038	2	US-08-971-217-1
C 17	35.4	3.5	2038	3	US-09-350-600-1
C 18	34	3.3	5084	1	US-08-306-691B-21
C 19	34	3.3	5084	5	PCT-US93-06251-25
C 20	33.2	3.3	15393	4	US-09-453-702B-191
C 21	32.2	3.2	453	4	US-09-484-970B-85
C 22	32	3.1	43360	4	US-09-453-702B-206
C 23	32	3.1	45325	4	US-09-453-702B-261
C 24	31.6	3.1	2703	4	US-09-482-273-75
C 25	31.6	3.1	2709	4	US-09-482-273-101
C 26	31.6	3.1	3001	4	US-09-539-333D-158
C 27	31.6	3.1	3001	4	US-09-539-333D-160

28 31.6 3.1 4203 2 US-08-866-757-1 Sequence 1, Appli
 29 31.6 3.1 4203 3 US-09-153-593-1 Sequence 1, Appli
 c 30 31.2 3.1 2559 4 US-09-569-098A-109 Sequence 109, App
 31 31 3.0 1221 4 US-09-134-001C-365 Sequence 365, App
 32 30.8 3.0 946 4 US-09-599-360B-42 Sequence 42, Appl
 c 33 30.6 3.0 684 4 US-09-134-001C-1999 Sequence 1999, Ap
 c 34 30.6 3.0 952 3 US-09-174-768-2 Sequence 2, Appli
 c 35 30.6 3.0 58407 4 US-08-916-421B-2 Sequence 2, Appli
 c 36 30.4 3.0 622 3 US-09-109-204-9 Sequence 9, Appli
 c 37 30.4 3.0 622 4 US-09-490-032-9 Sequence 9, Appli
 c 38 30.4 3.0 1422 4 US-09-227-357-18 Sequence 18, Appl
 c 39 30.4 3.0 1975 3 US-09-109-204-4 Sequence 4, Appli
 c 40 30.4 3.0 1975 4 US-09-490-032-4 Sequence 4, Appli
 c 41 30.4 3.0 3010 4 US-09-996-243-313 Sequence 313, App
 42 30.4 3.0 87350 3 US-08-781-891-79 Sequence 79, Appl
 43 30.4 3.0 87350 4 US-09-618-166-79 Sequence 79, Appl
 44 30.4 3.0 87543 4 US-09-791-211-3 Sequence 3, Appli
 45 30.2 3.0 13865 3 US-09-009-217-11 Sequence 11, Appli

ALIGNMENTS

RESULT 1

US-09-385-982-483/c
 ; Sequence 483, Application US/09385982
 ; Patent No. 6262334
 ; GENERAL INFORMATION:
 ; APPLICANT: ENDEGE, WILSON O., ET AL.
 ; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
 ; TITLE OF INVENTION: PRODUCTS: II
 ; FILE REFERENCE: CDMNA-260XX
 ; CURRENT APPLICATION NUMBER: US/09/385,982
 ; EARLIER FILING DATE: 1999-08-30
 ; EARLIER APPLICATION NUMBER: 09/328,111
 ; EARLIER FILING DATE: 1999-06-08
 ; EARLIER APPLICATION NUMBER: 60/117,393
 ; EARLIER FILING DATE: 1999-01-27
 ; EARLIER APPLICATION NUMBER: 60/098,639
 ; EARLIER FILING DATE: 1998-08-31
 ; NUMBER OF SEQ ID NOS: 544
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 483
 ; LENGTH: 605
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (1)..(605)
 ; OTHER INFORMATION: n = A,T,C or G
 ; US-09-385-982-483

Query Match 18.4%; Score 187.2; DB 3; Length 605;
 Best Local Similarity 65.5%; Pred. No. 3.2e-52;
 Matches 348; Conservative 0; Mismatches 158; Indels 25; Gaps 5;

QY 485 GTACATGGAGTATTCAACTGAATAAAGTGCACGGTAAAGCTTTTAAACGGTTAATT 544
 Db 597 GGACGAANGAANNAAGTCCCGGTAAGCTTTAAANGGTTAATTTTGTCAAATNCAGTA 538
 QY 545 TCTGTCAACAGTAG-ATGACAAATGCCCATCTTATCAGTGTCTCTTGAGCCCCCCT 603
 Db 537 GATAAANAANAAGTTTGTATACCAAGNAATGTTTCTTANGCTTTTCNTTNTCTT 478
 QY 604 TCCCTCTGTCTCCCTCCCGAGATGGCGGTTCAGTCCATATTATAACTGGCCATCTCA 663
 Db 477 AACACTGCCATGCNTCCCAATGSGGATTTAATCACTTAAACNGTGTCTGT 418
 QY 664 CAGTTGCTAACTTAGCAAGTGTCTTTTCTTAGACCCCTCTTAAACGACCAATATGCT 723
 Db 417 TAGTCGTAACCTTAGTAAGTGTCTTTTCTTATAGAACCCCTCTCTGACTGACCAATATGCT 358
 QY 724 GACCTGTACTATAAGATCTTCTGTATGATCGCATCTTTTGGTAGATAGTAGA 783

APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Utnes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericca C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US 08/449,315
FILING DATE: 24-MAY-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT 1989
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
APPLICATION NUMBER: US 07/769,122
FILING DATE: 27-SEP-1991
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:

NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/PL/CGC 1727
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2038 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLSCULE TYPE: cdna
FEATURE:
NAME/KEY: CDS
LOCATION: 932..1435
US-08-449-315-1
Query Match 3.5%; Score 35.4; DB 1; Length 2038;
Best Local Similarity 51.7%; Pred.No.0.27; 96; Indels 2; Gaps 1;
Matches 105; Conservative 0; Mismatches 96; Indels 2; Gaps 1;
QY 630 GGCGTTGAGTCATATTTAAACTGGCCATCCTCACAGTTGCTAAGTACGAGTCTTTT 689
Db 314 GACGTTAAATAACTATCAATTTGGACGTAAGAATATTTGGATATTTCTTCATTTCTTC 255
QY 690 CTTTAGGACC--CCCTTCTTAAGGACAAATATGTCGACCTGTACTATAAGATCTTTCTG 747
Db 254 TCTAATTTCTTCCCTTAATTTCCACITTGATGTTACTCATGTACTATCATATCTTTCTA 195
QY 748 ATAATGCATTCGGAGATTTTTTTTGGTAGATAGTAGAAGTGGCTTCTCTGTTTTCACCTTC 807
Db 194 TTATCTCATTCCAATCAGCGCTTGATCTCGAGTATAAATAATTTCTTTCTTTATCC 135
QY 808 TTTACTCAGCTGACTAGTCTTC 830
Db 134 AATCAGACTTGACTTTAGATTC 112
RESULT 6
US-08-444-803-1/c
Sequence 1, Application US/08444803
Patent No. 5654414
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Utnes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericca C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM: Floppy disk

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/165,667
;; FILING DATE: 8-MAR-1988
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/042,847
;; FILING DATE: 6-APR-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/632,441
;; FILING DATE: 21-DEC-1990
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/425,504
;; FILING DATE: 20-OCT-1989
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/848,506
;; FILING DATE: 6-MAR-1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/768,122
;; FILING DATE: 27-SEP-1991
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/580,431
;; FILING DATE: 7-SEP-1990
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/329,018
;; FILING DATE: 24-MAR-1989
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/045,957
;; FILING DATE: 12-APR-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Elmer, James Scott
;; REGISTRATION NUMBER: 36,129
;; REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (919)541-8614
;; TELEFAX: (919)541-8689
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2038 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 932..1435
;; US-08-449-043-1

Query Match 3.5%; Score 35.4; DB 1; Length 2038;
Best Local Similarity 51.7%; Pred.No. 0.27; 96; Indels 2; Gaps 1;
Matches 105; Conservative 0; Mismatches 96

QY	630	GGCGTTGAGTCCTATTTAAACTGGCCATCCTCACAGTTGCTAACTTAGCAAGTGGCTTTT	689
Db	314	GAGCTTAATAACTATCAATTTGGACGTAAGAATATTTGGATATTTCTTCAGTTCTTTC	255
QY	690	CTTTAGGACC--CCCTTCTTAAGGACATATGCTCGACCTGTACTATAGATCTTTCTG	747
Db	254	TCTAATTTCCCTCCCTTAATTTCCAACTTGATGTTACTCATGTAATATATCATATCTTCTA	195
QY	748	ATAATGCATTGGAGATTTTTTTGGTAGATAGAGAGCGTTCTCTGTTTTTCACTCTCC	807
Db	194	TTATCTCATTCCAATCAGCGCTGATCTGGAGTATATAAATTTCTATTTCTTAATCC	135
QY	808	TTTACTCAGTCACTAGTGTCTC	830
Db	134	AATCAGCACTGACTTTAGATTC	112

RESULT 8
US-08-456-265A-1/c
; Sequence 1, Application US/08456265A

;; Patent No. 5767369
;; GENERAL INFORMATION:
;; APPLICANT: Alexander, Danny C.
;; APPLICANT: Ryals, John A.
;; APPLICANT: Goodman, Robert M.
;; APPLICANT: Stinson, Jeffrey R.
;; TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
;; TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
;; NUMBER OF SEQUENCES: 111
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: CIBA-GEIGY Corporation
;; STREET: 520 White Plains Road, P.O. Box 2005
;; CITY: Tarrytown
;; STATE: New York
;; COUNTRY: USA
;; ZIP: 10591
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; OPERATING SYSTEM: IBM PC compatible
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/456,265A
;; FILING DATE: 31-MAY-95
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/181,271
;; FILING DATE: 13-JAN-1994
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/093,301
;; FILING DATE: 16-JUL-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/937,197
;; FILING DATE: 6-NOV-1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/678,378
;; FILING DATE: 1-APR-1991
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/305,566
;; FILING DATE: 6-FEB-1989
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/165,667
;; FILING DATE: 8-MAR-1988
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/042,847
;; FILING DATE: 6-APR-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/632,441
;; FILING DATE: 21-DEC-1990
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/425,504
;; FILING DATE: 20-OCT-1989
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/848,506
;; FILING DATE: 6-MAR-1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/768,122
;; FILING DATE: 27-SEP-1991
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/580,431
;; FILING DATE: 7-SEP-1990
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/368,672
;; FILING DATE: 20-JUN-1989
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/329,018
;; FILING DATE: 24-MAR-1989
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/045,957
;; FILING DATE: 12-APR-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Meigs, J. Timothy
;; REGISTRATION NUMBER: 38,241

REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727/DIV10

TELECOMMUNICATION INFORMATION:

TELEPHONE: (919)541-8587

TELEFAX: (919)541-8689

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2038 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cdna

FEATURE:

NAME/KEY: CDS

LOCATION: 932..1435

US-08-456-265A-1

Query Match 3.5%; Score 35.4; DB 1; Length 2038;

Best Local Similarity 51.7%; Pred. No. 0.27;

Matches 105; Conservative 0; Mismatches 96; Indels 2; Gaps 1;

QY 630 GCGCTTCAGTCCATATTAACTGGCCATCTCAGTGTCTAACTAGCAAGTCTTTT 689

Db 314 GAGCTTAAATCACTCAATCAATTTGGACGTAAAGAAATTTGGATATTCTTCAGTCTTTC 255

QY 690 CTTTAGGACC--CCCTTCTTAAACAGCAATATGCTGCACCTGTACTATAAGATCTTTCTG 747

Db 254 TCTAATTTCTTCCCTTAATTTCAATTTGAATGTTACTCAATGACTATCACTTTCTA 195

QY 748 ATAATGCAATCGGAGATTTTTTTTGGTAGATAGTAGAGTGGTTCCTGTTTTCACCTTCC 807

Db 194 TTATCTCATTCCAATCAGCGCTTGATCTGGAGTATAAATAATTTCTATTTTCTTATCCC 135

QY 808 TTACTCAGCTGACTAGTGTTC 830

Db 134 AATCAGCACTGACTTTAGATTC 112

RESULT 9

US-08-455-416-1/C

Sequence 1, Application US/08455416

Patent No. 5777200

GENERAL INFORMATION:

APPLICANT: Ryals, John A.

APPLICANT: Alexander, Danny C.

APPLICANT: Beck, James J.

APPLICANT: Duesing, John H.

APPLICANT: Friedrich, Leslie B.

APPLICANT: Goodman, Robert M.

APPLICANT: Harms, Christian

APPLICANT: Meins, Jr., Frederick

APPLICANT: Montoya, Alice

APPLICANT: Moyer, Mary B.

APPLICANT: Neuhaus, Jean-Marc

APPLICANT: Payne, George B.

APPLICANT: Sperison, Christoph

APPLICANT: Stinson, Jeffrey R.

APPLICANT: Uknes, Scott J.

APPLICANT: Ward, Eric R.

APPLICANT: Williams, Shericca C.

TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC

TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF

NUMBER OF SEQUENCES: 106

CORRESPONDENCE ADDRESS:

ADDRESSEE: CIBA-GEIGY Corporation

STREET: 7 Skyline Drive

CITY: Hawthorne

STATE: New York

COUNTRY: USA

ZIP: 10532

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA: US/08/455,416

FILING DATE: 31-MAY-1995

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/181,271

FILING DATE: 13-JAN-94

APPLICATION NUMBER: US 08/093,301

FILING DATE: 16-JUL-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/937,197

FILING DATE: 6-NOV-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/678,378

FILING DATE: 1-APR-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/305,566

FILING DATE: 6-FEB-1989

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/165,667

FILING DATE: 8-MAR-1988

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/042,847

FILING DATE: 6-APR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/632,441

FILING DATE: 21-DEC-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/425,504

FILING DATE: 20-OCT 1989

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/848,506

FILING DATE: 6-MAR-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/768,122

FILING DATE: 27-SEP-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/580,431

FILING DATE: 7-SEP-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/368,672

FILING DATE: 20-JUN-1989

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/329,018

FILING DATE: 24-MAR-1989

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/045,957

FILING DATE: 12-APR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Elmer, James Scott

REGISTRATION NUMBER: 36,129

REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727

TELECOMMUNICATION INFORMATION:

TELEPHONE: (919)541-8614

TELEFAX: (919)541-8689

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2038 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cdna

FEATURE:

NAME/KEY: CDS

LOCATION: 932..1435

US-08-455-416-1

Query Match 3.5%; Score 35.4; DB 1; Length 2038;

Best Local Similarity 51.7%; Pred. No. 0.27;

Matches 105; Conservative 0; Mismatches 96; Indels 2; Gaps 1;

QY 630 GCGCTTCAGTCCATATTAACTGGCCATCTCAGTGTCTAACTAGCAAGTCTTTT 689

Db 314 GACGTTAAATACTATCAATTTGGACGTAAGATATTTGGATATCTTCAGTTCTTTC 255
QY 690 CTTTAGGACC--CCCTTCTTAAGCAATATGCTGACCTGTACTATAAGATCTTTCTG 747
Db 254 TCTAATTTCTTCCCTTAATTTCCAACTTGTGATGTTACTCATGTACTATCATATTTCTA 195
QY 748 ATAATGCTATCGAGATTTTGGTAGATAGAGAGCGTCTCTGTTTTCACCTTCC 807
Db 194 TTATCTCATCAATCAACGGCTTGATCTGGAGTAAATAAATTTCTATTTCTTATCC 135
QY 808 TTACTCAGCTGACTAGTCTTC 830
Db 134 AATCAGACTTGACTTTAGATTC 112

RESULT 10

US-08-455-244-1/c
; Sequence 1, Application US/08455244
; Patent No. 5789214

GENERAL INFORMATION:

; APPLICANT: Ryals, John A.
; APPLICANT: Alexander, Danny C.
; APPLICANT: Beck, James J.
; APPLICANT: Duesing, John H.
; APPLICANT: Friedrich, Leslie B.
; APPLICANT: Goodman, Robert M.
; APPLICANT: Harms, Christian
; APPLICANT: Meins, Jr., Frederick
; APPLICANT: Montoya, Alice
; APPLICANT: Moyer, Mary B.
; APPLICANT: Neuhaus, Jean-Marc
; APPLICANT: Payne, George B.
; APPLICANT: Sperison, Christoph
; APPLICANT: Stinson, Jeffrey R.
; APPLICANT: Uknes, Scott J.
; APPLICANT: Ward, Eric R.
; APPLICANT: Williams, Shericca C.
; TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
; TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF

NUMBER OF SEQUENCES: 106

CORRESPONDENCE ADDRESS:

; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

FILING DATE: 31-MAY-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/181,271
; FILING DATE: 13-JAN-94
; APPLICATION NUMBER: US 08/093,301
; FILING DATE: 16-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/937,197
; FILING DATE: 6-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/678,378
; FILING DATE: 1-APR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/305,566
; FILING DATE: 6-FEB-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/165,667

; FILING DATE: 8-MAR-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/042,847
; FILING DATE: 6-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/632,441
; FILING DATE: 21-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/425,504
; FILING DATE: 20-OCT-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/848,506
; FILING DATE: 6-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/768,122
; FILING DATE: 27-SEP-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/580,431
; FILING DATE: 7-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/368,672
; FILING DATE: 20-JUN-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/329,018
; FILING DATE: 24-MAR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/045,957
; FILING DATE: 12-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919)541-8614
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2038 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 932..1435
; US-08-455-244-1

Query Match 3.5%; Score 35.4; DB 1; Length 2038;

Best Local Similarity 51.7%; Pred. No. 0.27;

Matches 105; Conservative 0; Mismatches 96; Indels 2; Gaps 1;

QY 630 GCGTTGAGTCCATATTTAACTGGCCATCCTCACAGTTGCTAACTAGCAAGTGTCTTT 689

Db 314 GAGCTTAAATAACTATCAAAATTTGGACGTAAGAATATTTGGATATCTTCAGTTCTTC 255

QY 690 CTTTAGGACC--CCCTTCTTAACGAGCAATATGCTGACCTGTACTATAAGATCTTCTG 747

Db 254 TCTAATTTCTTCCCTTAATTTCCACACTTGAATGTTACTCATGTACTATCATATTTCTA 195

QY 748 ATAATGCTATCGAGATTTTGGTAGATAGTAGAAGTGGGTCTCTGTTTTCACCTTCC 807

Db 194 TTATCTCATCAATCAACGGCTTGATCTGGAGTAAATAAATTTCTATTTCTTATCC 135

QY 808 TTACTCAGCTGACTAGTCTTC 830

Db 134 AATCAGACTTGACTTTAGATTC 112

RESULT 11

US-08-454-876-1/c
; Sequence 1, Application US/08454876
; Patent No. 5804693

GENERAL INFORMATION:

APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericca C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/454,876
FILING DATE: 31-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672

FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/Pl/CGC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
SEQUENCE CHARACTERISTICS:
SEQUENCE FOR SEQ ID NO: 1:
LENGTH: 2038 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 932..1435
US-08-454-876-1
Query Match 3.5%; Score 35.4; DB 1; Length 2038;
Best Local Similarity 51.7%; Pred. No. 0.27;
Matches 105; Conservative 0; Mismatches 96; Indels 2; Gaps 1;
QY 630 GGCGTTGAGTCCATATTTAACTGGCCATCCTCAGTTGCTAGTAAAGTCTTTT 689
Db 314 GACGTTAAATACTATCAAAATTTGGACGTAAGAATATTTGGATATTCTTCAGTTCTTTC 255
QY 690 CTTTAGGACC--CCCTTCTTAAAGAGCAATATGTCGACCTGTACTATAAGATCTTCTG 747
Db 254 TCTAATTTCTCTCCCTTAATTTCCAACTTGATGTTACTCATGTACTATCATCTTCTA 195
QY 748 ATAATGCATTGGAGATTTTTTTGGTAGATAGTAGAAGTGGTTCCTGTTTTCACCTTCC 807
Db 194 TTATCTCATTCCAATCAGCGCTTGATCTGGAGTATATAATAAATTTCTATTCTTATCCCC 135
QY 808 TTTACTCAGCTGACTAGTGCTTC 830
Db 134 AATCAGCACTTGACTTTAGATTC 112
RESULT 12
US-08-457-364-1/c
Sequence 1, Application US/08457364
Patent No. 5847258
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericca C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:

ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US 08/457,364
FILING DATE: 31-MAY-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/769,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-15825/P1/CGC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2038 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA

FEATURE:
NAME/KEY: CDS
LOCATION: 932..1435
US-08-457-364-1
Query Match 3.5%; Score 35.4; DB 2; Length 2038;
Best Local Similarity 51.7%; Pred. No. 0.27;
Matches 105; Conservative 0; Mismatches 96; Indels 2; Gaps 1;
QY 630 GCGTTGAGTCCATATTTAAACTGGCCATCTCCACAGTTGCTAACTTAGCAAGTGTCTTT 689
DB 3-4 GACGTTAAATAACTATCAAAATTTGACGTAAGAATATTTGGATATCTTCACTTTCTTC 255
QY 690 CTTTAGGACC--CCCTTCTTAACGAGCAATATGTCTGACCTGTACTATAAGATCTTTCTG 747
DB 254 TCTAATTTCTTCCCTTAATTTCCACACTTGATGTTACTCATGTACTATCATCTTTCTA 195
QY 748 ATAATGCATTCGGAGATTTTTTGGTAGATAGTAGAGTGCCTTCTGTGTTTTCACCTTC 807
DB 194 TTATCTCATTTCCAATCACGGCTTGATCTGGAGTATAAATAAATTTCTATTTCTTATCCC 135
QY 808 TTTACTCAGCTGACTAGTCTTC 830
DB 134 AATCAGACTTGACTTTAGATTC 112
RESULT 13
US-08-456-262-1/c
Sequence 1, Application US/08456262
Patent No. 5851766
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericca C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/456,262
FILING DATE: 31-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 1:
LENGTH: 2038 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 932..1435
US-08-456-262-1

Query Match 3.5%; Score 35.4; DB 2; Length 2038;
Best Local Similarity 51.7%; Pred.No. 0.27;
Matches 105; Conservative 0; Mismatches 96; Indels 2; Gaps 1;
QY 630 GCGCTTGAAGCCATATTAAACCTGGCCATCCTCACAGTTCTGTAACCTTAGCAAGTCTTTT 689
Db 314 GAGTTAAATACTATCAAAATTTGGAGTAAAGAAATATTTGGATATTCCTTCAGTCTTTC 255
QY 690 CTTTAGGACC--CCCTTCTTAAAGCAGCAATATGCTCACCTGTACTATAAGATCTTTCTG 747
Db 254 TCTAATTTCTCCCTTAATTTCAACTGTGATGTAATGTAATGTAATGTAATGTAATGTAAT 195
QY 748 ATAATGATTCGGAGATTTTTTTGGTAGATAGTAGAAGTGCCTTCCTGTTTTCACCTTCC 807
Db 194 TTATCTCATCCCAATCACCGGCTTGATCTGGAGTATAAATAAATTTCTATTTTCTATCCC 135

QY 808 TTTTACTAGCTGACTAGTGTTC 830
Db 134 AATCAGCTTGTGACTTTAGATTC 112
RESULT 14
US-08-456-240-1/C
; Sequence 1, Application US/08456240
; Patent No. 5856154
; GENERAL INFORMATION:
; APPLICANT: Rvals, John A.
; APPLICANT: Alexander, Danny C.
; APPLICANT: Beck, James J.
; APPLICANT: Duesing, John H.
; APPLICANT: Friedrich, Leslie B.
; APPLICANT: Goodman, Robert M.
; APPLICANT: Harms, Christian
; APPLICANT: Meins, Jr, Frederick
; APPLICANT: Montoya, Alice
; APPLICANT: Moyer, Mary B.
; APPLICANT: Neuhaus, Jean-Marc
; APPLICANT: Payne, George B.
; APPLICANT: Sperison, Christoph
; APPLICANT: Stinson, Jeffrey R.
; APPLICANT: Uknes, Scott J.
; APPLICANT: Ward, Eric R.
; APPLICANT: Williams, Shericca C.
; TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
; TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/456,240
; FILING DATE: 31-MAY-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/181,271
; FILING DATE: 13-JAN-94
; APPLICATION NUMBER: US 08/093,301
; FILING DATE: 16-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/937,197
; FILING DATE: 6-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/678,378
; FILING DATE: 1-APR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/305,566
; FILING DATE: 6-FEB-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/165,667
; FILING DATE: 8-MAR-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/042,847
; FILING DATE: 6-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/632,441
; FILING DATE: 21-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/425,504
; FILING DATE: 20-OCT-1989
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2038 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 932..1435
US-08-456-240-1

Query Match 3.5%; Score 35.4; DB 2; Length 2038;

Best Local Similarity 51.7%; Pred. No. 0.27; Mismatches 96; Indels 2; Gaps 1;
Matches 105; Conservative 0;

QY	630	GAGCTTGGTCCATATTTAACTGGCCATCCCTCACAGTTGCTAACTAGCAAGTGTTC	689
Db	314	GAGCTTAAATACTATCAATTTGGACGTAAGATATTTGGATATTTCTTCACTTCTTC	255
QY	690	CTTAGGACC--CCCTCTTAACGACCAATATCTCTGACCTGACTATAGATCTTTCG	747
Db	254	TCTAATTCCTCCCTTAATTTCCAACTTGATGTTACTCATGACTATCATACTTTCTA	195
QY	748	ATAATGCAATCGGAGATTTTTTTGGTAGATAGTAGAAGTGGCTTCCTGTTTTCACCTTC	807
Db	194	TTATCTCAATCCAACTACGCTTGATCTGGAGTATAATAAATTTCTATTTCTTATCCC	135
QY	808	TTTACTCAGCTGACTAGTGTTC	830
Db	134	AATCAGCACTGACTTTAGATTC	112

RESULT 15

US-08-455-736-1/c
Sequence 1, Application US/08455736
Patent No. 580328
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaus, Jean-Marc

APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Skinson, Jeffrey R.
APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericca C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/455,736
FILING DATE: 31-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-1994
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727

TELECOMMUNICATION INFORMATION:

TELEPHONE: (919)541-8614

TELEFAX: (919)541-8689

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2038 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 932..1435

US-08-455-736-1

Query Match

Best local Similarity 3.5%; Score 35.4; DB 2; Length 2038;

Matches 105; Conservative 0; Mismatches 96; Indels 2; Gaps 1;

QY 630 GGCGTTGAGTCCATATTTAACTGGCCATCCTCACAGTTGCTAACTTAGCAAGTCTTTT 689

Db 314 GAGTTAAATACTATCAAAATTTGGACGTAAAGAAATTTGGATATTCTTCAGTCTTTC 255

QY 690 CTTTAGGACC--CCCTTCTTAAAGAGCAATATGTCTGACCTGTACTATAAGATCTTTCTG 747

Db 254 TCTAATTTCTTCCCTTAATTTCCAACTTGATGTACTCATGTACTATCATCTTTCTA 195

QY 748 ATAATGCAITCGGAGATTTTGGTAGATAGTAGAGTGGTCCCTGTTTACCTTCC 807

Db 194 TTATCTCAITCCAATCACGGCTTGATCTGGAGTATAAATAATTTCTATTCTTATCC 135

QY 808 TTTACTCAGCTGACTAGTGCTTC 830

Db 134 AATCAGACTTGACTTAGATTC 112

Search completed: November 23, 2003, 15:53:44

Job time : 80.549 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 23, 2003, 14:02:08 ; Search time 360.418 Seconds
(without alignments)
9221.962 Million cell updates/sec

Title: US-09-717-321A-15

Perfect score: 1017

Sequence: 1 cccattctctgcagatt.....ccttgggtctgagggttc 1017

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2169961 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	290.6	28.6	1644	14	US-10-153-668-48
2	290.6	28.6	1646	14	Sequence 48, Appl
3	290.6	28.6	2051	9	US-10-153-668-46
4	290.6	28.6	2051	9	Sequence 46, Appl
5	285.2	28.0	1232	14	US-10-198-846-9874
6	283.2	27.8	2313	14	Sequence 9874, App
7	195.2	19.2	594	14	Sequence 1632, App
8	193.2	19.0	602	10	US-10-066-543-2864
9	193.2	19.0	602	10	Sequence 12919, A
10	193.2	19.0	602	13	US-09-878-178-1654
11	187.8	18.5	549	14	Sequence 1654, App
12	187.8	18.5	567	14	Sequence 1654, App
13	187.2	18.4	605	11	US-10-066-543-2378
14	185.4	18.2	477	14	Sequence 2878, App
15	172.2	16.9	432	10	US-10-066-543-3022
16	164	16.1	2167	12	Sequence 2964, App
					Sequence 483, App
					Sequence 3022, App
					Sequence 7612, App
					Sequence 500, App

17	147.4	14.5	409	14	US-10-066-543-13	Sequence 13, Appl
18	126.4	12.4	285	9	US-09-294-093B-506	Sequence 506, App
19	123.2	12.1	416	10	US-09-960-352-8781	Sequence 8781, App
20	108.8	10.7	293	14	US-10-066-543-3077	Sequence 3077, App
21	103.2	10.1	3740	11	US-09-764-891-9984	Sequence 9984, App
22	98.2	9.7	581	10	US-09-998-598-972	Sequence 972, App
23	91.6	9.0	565	10	US-09-998-598-296	Sequence 296, App
24	85.6	8.4	492	10	US-09-878-178-1503	Sequence 1503, App
25	85.6	8.4	492	13	US-10-046-935-1503	Sequence 1503, App
26	85.6	8.4	492	14	US-10-146-502-1503	Sequence 1503, App
27	82.2	8.1	398	11	US-09-918-995-34569	Sequence 34569, A
28	68	6.7	588	11	US-09-871-161-133	Sequence 133, App
29	62.2	6.1	426	10	US-09-960-352-7342	Sequence 7342, App
30	59	5.8	404	10	US-09-960-352-14206	Sequence 14206, A
31	57.6	5.7	1651	14	US-10-198-846-13019	Sequence 13019, A
32	52.4	5.2	162	10	US-09-920-300A-1549	Sequence 1549, App
33	52.4	5.2	162	12	US-10-099-926-1549	Sequence 1549, App
34	52.4	5.2	162	13	US-10-033-528-1549	Sequence 1549, App
35	40.8	4.0	289	9	US-09-923-876-6297	Sequence 6297, App
36	37.4	3.7	13712	12	US-10-311-455-1504	Sequence 1504, App
37	37.2	3.7	565	12	US-10-027-632-133365	Sequence 133365, A
38	37.2	3.7	565	13	US-10-027-632-133365	Sequence 133365, A
39	37	3.6	260209	12	US-10-025-966A-23	Sequence 23, Appl
40	37	3.6	260209	12	US-10-265-071-23	Sequence 23, Appl
41	36.6	3.6	1581	14	US-10-198-846-6480	Sequence 6480, App
42	36.4	3.6	725	12	US-10-027-632-11287	Sequence 11287, A
43	36.4	3.6	725	13	US-10-027-632-11287	Sequence 11287, A
44	36.2	3.6	968	12	US-10-027-632-121396	Sequence 121396, A
45	36.2	3.6	968	13	US-10-027-632-121396	Sequence 121396, A

ALIGNMENTS

RESULT 1

US-10-153-668-48
; Sequence 48, Application US/10153668
; Publication No. US20030092616A1
; GENERAL INFORMATION:
; APPLICANT: HONDA, Goichi
; APPLICANT: MATSUDA, Akio
; APPLICANT: MURAMATSU, Shuji
; APPLICANT: ISHIZAWA, Kenya
; TITLE OF INVENTION: STAT6 Activating Gene
; FILE REFERENCE: 1254-0207P
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: US/10/153,668
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/293,172
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/316,031
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/328,403
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: JP 2001-157043
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: JP 2001-260681
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: JP 2001-313175
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 488
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 48
; LENGTH: 1644
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (109)..(447)
US-10-153-668-48

Query Match 28.6%; Score 290.6; DB 14; Length 1644;
Best local Similarity 64.5%; Pred. No. 5.5e-83;
Matches 664; Conservative 0; Mismatches 304; Indels 61; Gaps 13;


```

QY 469 TAAATTTACACTACATGTGTACATGAGTAA---TTCACACTGAATAAAGTGTCAAGGTAA 525
Db 869 -ATTTTACACTACATGTGTACAGAAATGAAGTGTACCGGTAAAGTGTAAAGTGTAA 926
QY 526 AGCTTTTAAAGGTTAAATTTCTGTCAAAAGTAGATGACAAATGGCGATCTTATCAGTG 585
Db 927 ATTCTGTCAATGCGAGTAGATGATGAAGAAAGGTTGGTATTATCAGGAAATGTTTCT 986
QY 586 TCTCTCTTGAGCCCTTCCCTCTGTCTCCCTCCAGATGGGGGTTGATGCATAT 645
Db 987 TAAAGCTTTTCTTCTCTTACACCTGCGATGCTCCCAAAATGGGCATTTAATTCATCT 1046
QY 646 TTAAGCTGGCCATCTCACAGTTGCTAACTAGCAAGTGTCTTTCTTTAGGACCCCTTC 705
Db 1047 TTAAGCTGGTTGTTCTGTGTAGTGTCTTAACTAGTAAAGTGTCTTTCTTTATGAACCCCTTC 1106
QY 706 TTAAGGAGCAATATGCTGACCTGTACTATAAGATCTTTCTGATAATGCAATTCGGAGATT 765
Db 1107 TGACTGAGCAATATGCT-CTTGTATTAATAAATCTTTCTGATAATGCAATAGAGATT 1165
QY 766 TTTTGGTAGATAGTAGAGTGGTCTCTCTTTTACCTTCCCTTACTCAGCTGACTAGT 825
Db 1166 TTTTGGTAGATAGTAGAGTGGTCTCTCTTTTACCTTCCCTTACTCAGCTGACTAGT 1220
QY 826 GCTTCCCTTCTTCTTCTAGTAACTGGGTGTAGAAATCAAGTGTCTTCTTACAGTTT 885
Db 1221 GCTTCTTCTAGTTTCTAGTAACTAGTGTAAAGTGTCTTCTTACAGTTT 1280
QY 886 TAAACTATTTTAGATA-----TTCTGAAACATCACTGTTTCCGACAG 928
Db 1281 TAAATATTTTAGATAATCTTAACTATGAACCTTCTTAACTCACTGTTTCCAGAT 1340
QY 929 TACCACACTGTCTATGTATGTATGCGGCCCTCTAGACCTCACCCACGCGGACACATG 988
Db 1341 TACCACACTGTCTATGTATGTATGCGGCCCTCTAGACCTCACCCACGCGGACACAG 1399
QY 989 CTTCCGGTA 997
Db 1400 CCTCCTGTA 1408

```

RESULT 3

```

US-09-925-302-255
; Sequence 255, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 255
; LENGTH: 2051
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2)
; OTHER INFORMATION: n equals a.t.g, or c
; NAME/KEY: misc feature
; LOCATION: (50)
; OTHER INFORMATION: n equals a.t.g, or c
; NAME/KEY: misc feature
; LOCATION: (68)
; OTHER INFORMATION: n equals a.t.g, or c
; NAME/KEY: misc feature
; LOCATION: (2027)

```

```

; OTHER INFORMATION: n equals a.t.g, or c
; NAME/KEY: misc feature
; LOCATION: (2046)
; OTHER INFORMATION: n equals a.t.g, or c
US-09-925-302-255

Query Match      28.6%; Score 290.6; DB 9; Length 2051;
Best Local Similarity 64.5%; Pred. No. 6.5e-83;
Matches 664; Conservative 0; Mismatches 304; Indels 61; Gaps 13;

QY 1 CCCTATTCTTGTCTCAGATTAAAGATTGCAAAATACCTTGTCAACTAAAGTTG--GTT 57
Db 755 CCCCCCAATCTTGTTCAGATTAAAGTTTGCAAATATACCTTCTGAACCTACACTGCATTGTT 814
QY 58 GTCTGAGAACACGTAAGCACTAAGCTGTTGTAGAGACTTTGTCTTAAAGAGACTGCAGC 117
Db 815 GTCCGAGAACACCGGACACTGAATTTGCAAAAGACCTTCGCTTTGAGAAGACGGTAGC 874
QY 118 TTTCTGGCTCAGG-GGTGACAGCCCTCCGCTAGC-----TCCCAGACCGGTGACACA 169
Db 875 TTTCTGAGTTAGAGGTGACAGACACTTGTCTCTCTATGTAGTTCTCAGATGCGTAAGCA 934
QY 170 GCACAGCCTCTTAAATGACACACGCTGCCATGTAAAGCACTGTAACTATCAGCCCATGCT 229
Db 935 GAACAGCCTCCCGAATGAAGCGTTGCCATTGAACTCAACAGTGAAGTAGCAGCACGTT 994
QY 230 CATTACGTAACCTTTGACTGTACGTCAAGATGGGTGTAAAGCTCTGCTCTTTGATTTC 289
Db 995 CCGGACATAAATGTTACTGTAATGAGTGAGCGTAGCAGCTCAGCTCTTTGGATCAGTC 1054
QY 290 TAGTGAGTTCTCTAAATACCAGCTGACCGGCTTCTGCAGGCTTTGAACAGAACTCTGGC 349
Db 1055 TTTGTGATTTCAATAGCGAGTTTCTGACGAGCTTTTCCGAGAGATTTTGAACAGAACTG-- 1112
QY 350 TCCTGTGTTGCTCTTAAACGAAATATTCTGTCTCTAGTCGTTGGGTGTCTGGGTGGAGTGT 409
Db 1113 ----CTATTTCCTCTAATGAAGATTCTGTT--TAGCTGGGTGTGCGGGTGGGTGT 1166
QY 410 GTGAAACACAGAGCTCATCAAGGAGACAGACATATTTTGAC-TAATATGAAGTAGAGAT 468
Db 1167 GT-----GTGATCAAAGGACAAAGACAGATATTTTGACAAATACGAAAGTGAG-- 1214
QY 469 TAAATTTACACTACATGTGTACATGGAGTAA---TTCACACTGAATAAAGTGTCAAGGTAA 525
Db 1215 -ATTTTACACTACATGTGTACAGGAATGAAGTGTCAAGGTAAAGTGTCTTAAAGGTAA 1272
QY 526 AGCTTTTAAAGGTTAAATTTCTCTCAACAGTAGATGACAAATGGCGATCTTATCAGTG 585
Db 1273 ATTTCTGTCAATGCGAGTAGATGATGAAGAAAGTGGTATTATCAGGAAATGTTTCT 1332
QY 586 TCTCTCTTGAGCCCTTCCCTCTGTCTCCCTCCAGATGGGGGTTGAGTCCATAT 645
Db 1333 TAAAGCTTTTCTTCTCTTACACCTGCCATGCTCCCAAAATTTGGCATTTAATTCATCT 1392
QY 646 TTAAGCTGGCCATCTCTCAGAGTGTCTTACAGTGTCTTCTTTAGGACCCCTTC 705
Db 1393 TTAAGCTGGTTGTTCTGTGTAGTGTCTTAACTTAGTAAAGTCTTTTCTTATAGAACCCCTTC 1452
QY 706 TTAACGAGCAATATGTCGACCTGTACTATAAGATCTTTCTGATAATGCAATCGGAGATT 765
Db 1453 TGACTGAGCAATATGCT-CTTGTATTATAAATCTTTCTGATAATGCAATGAGAAGTT 1511
QY 766 TTTTGGTAGATAGTAGAAGTGGTTCCTGTTTCTACCTTCCCTTTACTCAGCTGACTAGT 825
Db 1512 TTTTGTGCTAGTAGTAAAGTGTCTTCCATGTTAC-----TTTATTGAGAGCTAATAAGT 1566
QY 826 GCTTCCCTTCTGTTTCTAGTAACTGGGTGTAGAAATCAGTGTCTGCGGCTTTACAGTTT 885
Db 1567 GCTTCTCTTAGTATTTCTAGTAACTAGGTGTAAAGTGTCTGTTGACGCTTTATAGTTT 1626
QY 886 TAAACTATTTTAGATA-----TTCTGAAACATCACTGCTTCTGACAG 928
Db 1627 TAAATATTTTAGATAATTTCTTAAACTATGAACCTTCTTAAACATCACTGTTCTTCCAGAT 1686

```



```

QY 65 GAACACGTAAGCACTAAGCTGTTGAGAGACTTTGTCCTTAAGAGACTGCAGCTTCTGGG 124
Db 61 GAAACACCGAGCACTGAACCTTAGCAAGACCTTGTCTTTGAGAGACGGTAGCTTCTGCA 120
QY 125 CTCAGG-GGTGACAGCCTCCCGTAGC-----TCCGAGACCGGTGTGACACAGCACAGC 176
Db 121 GTTAGGAGGTGCAGACACTTGTCTCTCCTATGTAGTTCTCAGATCGGTAAAGCAGACAGC 180
QY 177 CTCCTTAATGACAGCTGCCATGTAACGACCTGTAACTTATCAGCCCATGCTCATTAAG 236
Db 181 CTCCTTAATGAGCGTTGCCATTAAGCTTAACTTAACTTAACTTAACTTAACTTAACTTAA 240
QY 237 TAACCTTTGACTGTACGTACAGATGGGTGTAAACAGCTGTCTCTTTGATTTCAATGAG 296
Db 241 TAACATTTGACTGTAAAGGAGTAGCTAGACGTCTTGGATCAGCTTTGAGTCTTTGTA 300
QY 297 TTTCTCTAAATACACAGTGACCGGCTTTCTGACGCTTTTGAACAGAACTCTGCTCCTGTG 356
Db 301 TTTCTATAGCGAGTTTCTGACCACTTTTGGGAGATTTTGAACAGAACTG-----CTA 354
QY 357 TTGCTCTAAGAGATTTCTGTTCTTAGTGTGGGTGTGCTGGGTGAGTGTGAAAC 416
Db 355 TTTCTCTAATGAAGAAATCTGTGTT--TAGCTGTGGGTGTGCGCGGTGGGTGTGTT---- 407
QY 417 ACGACGTCATCAAGAGGAGACAGACAGTATTTTGAC-TAATATGAAGTAGAGATTAATTTA 475
Db 408 -----GTGATCAAGAGCAAGACAGATTTTGAACAAATACGAAGTGGAG-----ATT 458
QY 476 CACTACATTTGATCATGGAGT---AAATCAACTGAATAAAGTGTACCGGTGAAGCTTTT 532
Db 459 CACTACATTTGATCAAGGAATGAAAGTGTACCGGTGAAGAACTCTAAAGGTTAATTTCTG 518
QY 533 TAACGGTTAATTTCTGTCACACAGTAGATGACAAATGGCCGATCTTATCAGTGTCTCTCT 592
Db 519 TCAATGCAATGATGATGAAGAAAGAGGTGGTATTATCAGGAATGTTTCTTAAAGCTT 578
QY 593 TGAGGCCCCCTTCCCGCTGTGCTCCCTCCAGATGGGGTGTGAGTCCATATTTAACT 652
Db 579 TTCTCTTCTTACACTGCGATGCTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCC 638
QY 653 GGCCATCTCACAAGTTGCTAACTTAGCAAGTGTCTTTCTTTAGGACCCCTTCTTAAACGA 712
Db 639 GGTGTGTTCTGTAGTGTGCTAACTTAGTGTGCTTTCTTTATAGAACCCCTTCTGACTGA 698
QY 713 GCAATATGCTGACCTGACTATAAGATCTTCTGATATGATGATGATGATGATGATGATG 772
Db 699 GCAATATGCTGCTTCTGATATTAATAATCTTCTGATATGATGATGATGATGATGATGAT 757
QY 773 TAGATAGTAGAAGTGGTCTCTGTTTCACTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTT 832
Db 758 CGATTAGTAAAGTGTCTTCCATGTAC-----TTTATTCAGAGCTAAATAGTGTCTTCC 812
QY 833 TTGCTTTCTAGTAAGTGGGTGAGAAATCACGTGCTGCGGCTTTTACAGTTTAACTA 892
Db 813 TTAGTTTCTAGTAAGTGGGTGAGAAATCACGTGCTGCGGCTTTTATAGTTTAACTA 872
QY 893 TTTTAGATA-----TTCTGAACATCTCTGCTTGGCCAGAGTACCAAC 935
Db 873 TTTTAGATAATTTTAACTATGAACCTTCTTAACTATCTTCTTCTTCTTCTTCTTCTTCTT 932
QY 936 ACTGTCATGTGATGATGCGGCCCTCTAGACCTTCAACCGCGACAGCATGCTTCCGG 995
Db 933 ACTGTCATGTGACCAATAGT-ACCTCTTTTACCTCGCCACGCGACACAGCCTCCTG 991
QY 996 TA 997
Db 992 TA 993

```

RESULT 6
 US-10-198-846-12919
 ; Sequence 12919, Application US/10198846

```

; Publication No. US20030099974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10198,846
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 12919
; LENGTH: 2313
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-198-846-12919

```

```

Query Match      27.8%   Score 283.2; DB 14; Length 2313;
Best Local Similarity 63.5%; Pred. No. 1.8e-80;
Matches 65; Conservative 0; Mismatches 318; Indels 57; Gaps 12;

```

```

QY 1 CCCTATTTCTGTCTGATTAAGAAATTCGCAAAATACCTTGTGAACCTAAGTTGC---GTT 57
Db 1224 CCCCCATTTCTGTCTGATTAAGAAATTCGCAAAATACCTTGTGAACCTAAGTTGC---GTT 1293
QY 58 GTGCTGAGACAGTAAGCACTAAGCTGTTGAGAGACTTTGCTTAAAGAGACTGACG 117
Db 1284 GTCCGAGAACACCGGACACTGAACCTTTCGCAAGACCTTCTGAGAGACGCTAGC 1343
QY 118 TTCTGGGCTCAGG-GGTGACAGCCCTCCGTCAGC-----TCCGACCGTGTGACACA 169
Db 1344 TTCTGAGTTAGAGGTGCGACACTTGTCTCTCTATGATGTTCTCAGATGCGTAAAGCA 1403
QY 170 GCACAGCTCTTTAATGACACGCTGCGCATGTAACGACCTGTAACTTATCAGCCCATGCT 229
Db 1404 GAACAGCTCCCGAATGAAGCGTTGCCATTTGAACCTCAGCTGATTTAGCAGACGCTGT 1463
QY 230 CATTAGTAAGTCTTGTGATGCTGCTGACGATGGGTGTAAAGCTGCTCTTTGATTTCA 289
Db 1464 CCGACATAAACAATGATGCTGATGAGTGTGAGCTGAGCTCAGCTCTTTGATCAGTC 1523
QY 290 TAGTGAGTTCTCTAAATAACAGCTGACCGCTTCTGACGCTTTTGAACAGAACTCTGGC 349
Db 1524 TTGTGATTTTATGAGGAGTTTCTGACCACTTTTTCGGGAGATTTTGAACAGAA--- 1578
QY 350 TCTGTGTTGCTCTAACGAAGTATTTCTGTTCTCTAGTGTGCTGCTGGGTGGAGTGT 409
Db 1579 --CTGTATTTCTCTAATGAAGAAATTTCTGT--TAGCTTGTGCTGCTGGGTGGAGTGT 1634
QY 410 GTGAACACAGCTGATCAAGAGAGACAGAGTATTTTGAC-TAATATGAAGTAGAGAT 468
Db 1635 GT-----GTGATCAAGAGCAAAAGACAGTATTTTGAACAAATACGAAGTGTGAGAT 1684
QY 469 TAATTTACACTACATTTGATGAGTAAATTCACATGAATAAAGTGTACCGGTAAAGC 528
Db 1685 TACACTACATTTGATCAAGAAATGAA--AGTGTACGGGTAAAGAACTCTAAAGAGTTAAT 1742
QY 529 TTTTAAACGTTTAAATTTCTGTCAAAACAGTAGATGACAAATGCGCGATTTTATCAGTGTCT 588
Db 1743 TCTGTCAATGAGTAGATGATGAAGAAAGTGTGATTTATCAGGAATGTTTCTTAA 1802
QY 589 CTCCTTGAAGCCCCCTTCCCTCTGCTGCTCCCTCCAGATGGGCGGTGAGTCAATATTA 648
Db 1803 GCTTTTCTCTTTTACACTGCTGCTCCCTCCCAAAATTTGGCATTTTAAATCACTTTA 1862
QY 649 AACTGGCATCTCAGTTGCTAAGTGTAGCAAGTGTCTTTCTTTAGGACCCCTTCTTA 708
Db 1863 AACTGGTGTCTGTAGTGTGCTAAGTGTCTTTCTTTATAGAACCCCTTCTGA 1922

```

QY 709 ACGAGCAATATGCTGACCTGACTACTATAAGATCTTTCTGATAATGCAATTCGGAGATTTT 768
 Db 1923 CTGAGCAATATGCTT-CCTTGATATATAAAATCTTTCTGATAATGCAATTAGAAGGTTTTT 1981
 QY 769 TTGGTAGATAGTGAAGTGGCTTCTGTTTTCACCTTCTTCTTACTCAGCTGACTAGTGTCT 828
 Db 1982 TTGTGATAGTGAAGTGGCTTCTGTTTTCACCTTCTTCTTACTCAGCTGACTAGTGTCT 2036
 QY 829 TCCCTTCTGTTTCTAGTAAGTGGTGTAGAAATACAGTGTCTGGGCTTTTACAGTTTTTAA 888
 Db 2037 TTCTTAGTCTTCTAGTAAGTGGTGTAGAAATACAGTGTCTGGGCTTTTACAGTTTTTAA 2096
 QY 889 ACTATTTTAGATA-----TTCTGAACAATACATGCTTGTGCGGAGTAC 931
 Db 2097 AATATTTTAGATAATCTTAAACTATGAACCTTCTTAAATACATGCTTGTGCGGAGTAC 2156
 QY 932 CAACACTGCTGATGATGATGCGGCTCTAGACCTCACCACGCGGACATGCTT 991
 Db 2157 CGACACTGCTGATGATGCGGCTCTAGACCTCACCACGCGGACATGCTT 2215
 QY 992 CCGGTA 997
 Db 2216 CCGTGA 2221

RESULT 7
 US-10-066-543-2864
 ; Sequence 2864, Application US/10066543
 ; Publication No. US20030087818A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Jiang, Yuqiu
 ; APPLICANT: Pyle, Ruth A.
 ; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Indrias, Carol Yoseph
 ; APPLICANT: Lodes, Michael J.
 ; APPLICANT: Secrist, Heather
 ; APPLICANT: Carter, Darrick
 ; APPLICANT: Fanger, Gary R.
 ; APPLICANT: Smith, Carole L.
 ; APPLICANT: Durham, Margarita
 ; APPLICANT: Stolk, John A.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 ; FILE REFERENCE: 210121.563
 ; CURRENT APPLICATION NUMBER: US/10/066,543
 ; CURRENT FILING DATE: 2002-01-31
 ; NUMBER OF SEQ ID NOS: 3417
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 2864
 ; LENGTH: 594
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-066-543-2864

Query Match 19.2%; Score 195.2; DB 14; Length 594;
 Best Local Similarity 70.9%; Pred. No. 2.7e-52;
 Matches 373; Conservative 0; Mismatches 113; Indels 40; Gaps 7;
 QY 505 GAATAAAGTGTACGGGTAAAGCTTTTAAACGGTTAAATTTCTGTCAAA--CAGTAGATG 562
 Db 6 GAATGAAGTGTACGGGTAAAGCTTTTAAACGGTTAAATTTCTGTCAAA--CAGTAGATG 65
 QY 563 ACAAA-----GGCGATCTTATCAGTGTCTCT-----CTTGAGCCCCCTTCCC 608
 Db 66 ATGAAGAAAGGTGTGATATATCAGGAATGTTTCTTAAAGCTTTTCTTCTTCTTACAC 125
 QY 609 CTGCTGTCCCTCCCGATGGGCTGTAGTCCATATTTAACTGGCCATCTTCACAGTT 668
 Db 126 CTGCGATGCTTCCCGAATGGGCAATTAATCATCTTTAACTGGTGTCTGTAGTC 185
 QY 669 GCTAACTTAGCAAGTGTCTTTCTTTAGGACCCCTTCTTAAACGAGCAATATGCTGACCT 728

Db 186 GCTAACTTAGTAAGTGTCTTTTCTTATAGAACCCCTTCTGACTGAGCAATATGCTT-CCTT 244
 QY 729 GTACTATAAGATCTTTCTGATAATGCAATTCGGAGATTTTCTGATAATGCAATTCGGAGTGC 788
 Db 245 GTATTATAAATCTTTCTGATAATGCAATTCGGAGATTTTCTGATAATGCAATTCGGAGTGC 304
 QY 789 GTTCTGTTTTTCACTTCTTCTTACTCAGCTGACTAGTGTCTTCCCTTCTGTTTTCTAGTAAC 848
 Db 305 TTTCCATGTTAC-----TTTATTAGAGCTAATAAGTGTCTTCTTCTTCTAGTAAC 359
 QY 849 TGGGTGTAGAAATACAGTGTCTGGGCTTTACAGTTTTTAACTATTTTAGATA-----901
 Db 360 TAGGTGTAAATCATGTGTGTCAGCTTTATAGTTTTTAAATATTTTAGATAATTTCTTA 419
 QY 902 -----TTCTGAACAATACATGCTTGTGCGGAGTACCAACACTGTGCTGATGTA 951
 Db 420 AACTATGAACCTTCTTAAATCATGCTTGTGCGGAGTACCAACACTGTGCTGATGTA 479
 QY 952 TGCCGCCCTCTCTAGACCTCACCACGCGGACATGCTTCCGTA 997
 Db 480 TACTG-ACCTCTTTTACCTGCGCCACGCGGACACACGCTCCTGTA 524

RESULT 8
 US-09-878-178-1654/c
 ; Sequence 1654, Application US/09878178
 ; Patent No. US20020177552A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Jiang, Yuqiu
 ; APPLICANT: Harlocker, Susan L.
 ; APPLICANT: Secrist, Heather
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 ; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
 ; FILE REFERENCE: 210121.527
 ; CURRENT APPLICATION NUMBER: US/09/878,178
 ; CURRENT FILING DATE: 2001-06-08
 ; NUMBER OF SEQ ID NOS: 2237
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1654
 ; LENGTH: 602
 ; TYPE: DNA
 ; ORGANISM: Homo sapien
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (1)-(602)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-878-178-1654

Query Match 19.0%; Score 193.2; DB 10; Length 602;
 Best Local Similarity 70.5%; Pred. No. 1.2e-51;
 Matches 371; Conservative 0; Mismatches 115; Indels 40; Gaps 7;
 QY 505 GAATAAAGTGTACGGGTAAAGCTTTTAAACGGTTAAATTTCTGTCAAA--CAGTAGATG 562
 Db 592 GAATGAAGTGTACGGGTAAAGCTTTTAAACGGTTAAATTTCTGTCAAA--CAGTAGATG 533
 QY 563 ACAAA-----GGCGATCTTATCAGTGTCTCT-----CTTGAGCCCCCTTCCC 608
 Db 532 ATGAAGAAAGGTGTGATATATCAGGAATGTTTCTTAAAGCTTTTCTTCTTCTTACAC 473
 QY 609 CTGCTGTCCCTCCCGATGGGCTGTAGTCCATATTTAACTGGCCATCTTCACAGTT 668
 Db 472 CTGCCATGCTTCCCGAATGGGCAITTAATCATCTTTAACTGGTGTCTGTAGTC 413
 QY 669 GCTAACTTAGCAAGTGTCTTTCTTTAGGACCCCTTCTTAAACGAGCAATATGCTGACCT 728
 Db 412 GCTAACTTAGTAAGTGTCTTTCTTATAGAACCCCTTCTGACTGAGCAATATGCTT-CCTT 354
 QY 729 GTACTATAAGATCTTTCTGATAATGCAATTCGGAGATTTTCTGATAATGCAATTCGGAGTGC 788
 Db 353 GTATTATAAATCTTTCTGATAATGCAATTCGGAGATTTTCTGATAATGCAATTCGGAGTGC 294
 QY 789 GTTCTGTTTTTCACTTCTTCTTACTAGTGTCTTCCCTTCTGTTTTCTAGTAAC 848

```

Db      293  TTTCAGTTAC-----TTTATTAGAGCTAATAGTCTTCTTCTTAGTTTTCAGTAAC 239
QY      849  TGGGTAGAAATCAGCTGTGCGGCTTTACAGTTTTTAACTATTTAGATA----- 901
Db      238  TAGGTGTAATAATCATGTGTGACGCTTTATAGTTTTTAAATATTTAGATAATTTCTTA 179
QY      902  -----TTCTGAACATCACTGTCTGCCAGAGTACCAACACATGTCATGTGATGA 951
Db      178  AACTATGAACCTTTCTTAACATCACTGTCTTGCCAGATTACCGACACTGGCACATTGACCAA 119
QY      952  TGGCGCCCTCTAGACCTCACCACCGGACACATGCTTCCGGTA 997
Db      118  TACTG-ACCCTCTTTACTCGCCACCGGACACACGCTCTCTGTA 74

```

RESULT 9

```

US-10-046-935-1654/c
; Sequence 1654, Application US/10046935
; Publication No. US20020156011A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Jiang, Yugu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; APPLICANT: Wang, Aijun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.527C1
; CURRENT APPLICATION NUMBER: US/10/046,935
; NUMBER OF SEQ ID NOS: 2239
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1654
; LENGTH: 602
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 569_578
; OTHER INFORMATION: n = A,T,C or G

```

```

US-10-046-935-1654

```

```

Query Match      19.0%; Score 193.2; DB 13; Length 602;
Best Local Similarity 70.5%; Pred. No. 1.2e-51;
Matches 371; Conservative 0; Mismatches 115; Indels 40; Gaps 7;

QY      505  GAATAAAGTGTACGGGTAAAGCTTTTAAACGGTTAAATTTCTGTCAAA--CAGTAGATG 562
Db      592  GAATGAAGTGTACGGGTAAAGCTTTAAAGGTTAAATTTCTGTCAATGCGAGTAGATG 533
QY      563  ACAAAT-----GGCGATCTTATCAGTGTCTCT-----CTTGAGCCCCCTTCCCC 608
Db      532  ATGAAGAAGAGTGTGTTATTAATCAGGAATGTTTTCTTAAGCTTTTCTTCTTTACAC 473
QY      609  CTGCTCTCCCTCCCGAGATGGGCGTTGAGTCCATATTTAAACTGGCCATCCTCACAGTT 668
Db      472  CTGCCATGCTCCCAATGGGCATTTAATTCATCTTTAACTGTTGTTCTGTGTAGTC 413
QY      669  GCTAATAGAGTGTCTTCTTAAAGTGTGAGTCCATATTTAACTGGCCATCCTCACAGTT 668
Db      412  GCTAATAGTGTCTTCTTAAAGTGTGAGTCCATATTTAACTGTTGTTCTGTGTAGTC 413
QY      729  GTACTATAGAGTGTCTTCTTAAAGTGTGAGTCCATATTTAACTGGCCATCCTCACAGTT 728
Db      412  GCTAATAGTGTCTTCTTAAAGTGTGAGTCCATATTTAACTGTTGTTCTGTGTAGTC 354
QY      729  GTACTATAGAGTGTCTTCTTAAAGTGTGAGTCCATATTTAACTGGCCATCCTCACAGTT 728
Db      353  GTATTATAAATCTTCTTAAAGTGTGAGTCCATATTTAACTGTTGTTCTGTGTAGTC 294
QY      789  GTTCCATGTTTACCTTCTTAAAGTGTGAGTCCATATTTAACTGGCCATCCTCACAGTT 848
Db      293  TTTCATGTTAC-----TTTATTAGAGCTAATAGTCTTCTTCTTAGTTTCTAGTAAC 239
QY      849  TGGGTAGAAATCAGCTGTGCGGCTTTACAGTTTTTAACTATTTAGATA----- 901
Db      238  TAGGTGTAATAATCATGTGTGACGCTTTATAGTTTTTAAATATTTAGATAATTTCTTA 179
QY      902  -----TTCTGAACATCACTGTCTTGCCAGAGTACCAACACATGTCATGTGATGA 951
Db      178  AACTATGAACCTTTCTTAACATCACTGTCTTGCCAGATTACCGACACTGGCACATTGACCAA 119

```

```

Db      238  TAGGTGTAATAATCATGTGTGACGCTTTATAGTTTTTAAATATTTAGATAATTTCTTA 179
QY      902  -----TTCTGAACATCACTGTCTTGCCAGAGTACCAACACATGTCATGTGATGA 951
Db      178  AACTATGAACCTTTCTTAACATCACTGTCTTGCCAGATTACCGACACTGGCACATTGACCAA 119
QY      952  TGGCGCCCTCTAGACCTCACCACCGGACACATGCTTCCGGTA 997
Db      118  TACTG-ACCCTCTTTACTCGCCACCGGACACACGCTCTCTGTA 74

```

RESULT 10

```

US-10-146-502-1654/c
; Sequence 1654, Application US/10146502
; Publication No. US20030069180A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Jiang, Yugu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; APPLICANT: Wang, Aijun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.527C2
; CURRENT APPLICATION NUMBER: US/10/146,502
; CURRENT FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2241
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1654
; LENGTH: 602
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 569_578
; OTHER INFORMATION: n = A,T,C or G

```

```

US-10-146-502-1654

```

```

Query Match      19.0%; Score 193.2; DB 14; Length 602;
Best Local Similarity 70.5%; Pred. No. 1.2e-51;
Matches 371; Conservative 0; Mismatches 115; Indels 40; Gaps 7;

QY      505  GAATAAAGTGTACGGGTAAAGCTTTTAAACGGTTAAATTTCTGTCAAA--CAGTAGATG 562
Db      592  GAATGAAGTGTACGGGTAAAGCTTTAAAGGTTAAATTTCTGTCAATGCGAGTAGATG 533
QY      563  ACAAAT-----GGCGATCTTATCAGTGTCTCT-----CTTGAGCCCCCTTCCCC 608
Db      532  ATGAAGAAGAGTGTGTTATTAATCAGGAATGTTTTCTTAAGCTTTTCTTCTTTACAC 473
QY      609  CTGCTCTCCCTCCCGAGATGGGCGTTGAGTCCATATTTAAACTGGCCATCCTCACAGTT 668
Db      472  CTGCCATGCTCCCAATGGGCATTTAATTCATCTTTAACTGTTGTTCTGTGTAGTC 413
QY      669  GCTAATAGAGTGTCTTCTTAAAGTGTGAGTCCATATTTAACTGGCCATCCTCACAGTT 728
Db      412  GCTAATAGTGTCTTCTTAAAGTGTGAGTCCATATTTAACTGTTGTTCTGTGTAGTC 354
QY      729  GTACTATAGAGTGTCTTCTTAAAGTGTGAGTCCATATTTAACTGGCCATCCTCACAGTT 728
Db      353  GTATTATAAATCTTCTTAAAGTGTGAGTCCATATTTAACTGTTGTTCTGTGTAGTC 294
QY      789  GTTCCATGTTTACCTTCTTAAAGTGTGAGTCCATATTTAACTGGCCATCCTCACAGTT 848
Db      293  TTTCATGTTAC-----TTTATTAGAGCTAATAGTCTTCTTCTTAGTTTCTAGTAAC 239
QY      849  TGGGTAGAAATCAGCTGTGCGGCTTTACAGTTTTTAACTATTTAGATA----- 901
Db      238  TAGGTGTAATAATCATGTGTGACGCTTTATAGTTTTTAAATATTTAGATAATTTCTTA 179
QY      902  -----TTCTGAACATCACTGTCTTGCCAGAGTACCAACACATGTCATGTGATGA 951
Db      178  AACTATGAACCTTTCTTAACATCACTGTCTTGCCAGATTACCGACACTGGCACATTGACCAA 119

```

QY 952 TGCGCCCGCTTAGACCTCCACCGGGACACATGCTTCGGTA 997
DB 118 TACTG-ACCCTCTTTACCTCGCCCGGGACACACGCTCTCTGTA 74

RESULT 11

US-10-066-543-2878/c
; Sequence 2878, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:

; APPLICANT: Jiang, Yugu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indrias, Carol Yoseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Margarita
; APPLICANT: Stolk, John A.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.563

; CURRENT APPLICATION NUMBER: US/10/066,543
; CURRENT FILING DATE: 2002-01-31

; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2878

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc.feature

; LOCATION: 526

; OTHER INFORMATION: n = A,T,C or G

US-10-066-543-2878

Query Match 18.5%; Score 187.8; DB 14; Length 549;
Best Local Similarity 73.3%; Pred. No. 6.4e-50;
Matches 304; Conservative 0; Mismatches 87; Indels 24; Gaps 4;

QY 600 CCCTTCCCGCTGTGCTCCCTCCCGATGGGGGTTGAGTCCATATTTAACTGGCCATC 659
DB 481 CTCCTACCTGCGATGCTCCCTCCCAATTTGGGCAATTAATCACTTTAACTGGTTGT 422
QY 660 CTCACAGTTGCTAACTAGCAAGTGTCTTTCTTTAGGACCCCTCTCTTAACGAGCAATAT 719
DB 421 CTGTTAGTCGTAACCTAGTAAGTGTCTTTCTTTATAGAACCCCTCTGACTGAGCAATAT 362
QY 720 GTCTGACCTGTACTATAAGATCTTTCTGATAATGCAATCGGAGATTTTTTGGTAGATAG 779
DB 361 GCCT-CCCTGTATTATAAATCTTTCTGATAATGCAATCGGAGATTTTTTGGTAGATAG 303
QY 780 TAGAAGTGGTTCCTGTTTTCACCTTCTTACTCAGCTGACTAGTGTCTTCCCTTCGTTT 839
DB 302 TAAAGTGTCTTCCATGTTAC-----TTTATTACAGAGCTAATAAGTGTCTTCCCTTAGTT 248
QY 840 TCTAGTAAGTGGGTGTAGAAATCACTGTCTGGGGTTTACAGTTTTTAACTATTTTGA 899
DB 247 TCTAGTAACCTAGGTGTAAATAATCATGTGTGACGCTTTATAGTTTTTAAATAATTTTGA 188
QY 900 TA-----TTCTGAACATCACTGTCTTCCAGAGTACCAACACTGTCA 942
DB 187 TAATCTTAAATATGAACCTCTTTAACTATCACTGTCTTCCAGATTTACGACACTGTCA 128
QY 943 TGTGATTGATGGCGCCCTCTAGACCTCACTACCGGAGACACATGCTTCCGGTA 997
DB 127 CTTGACCAATACCTG-ACCCTCTTTACCTCGCCCGGGACACACGCTCTCTGTA 74

RESULT 12

US-10-066-543-2964/c
; Sequence 2964, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:

; APPLICANT: Jiang, Yugu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indrias, Carol Yoseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Margarita
; APPLICANT: Stolk, John A.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.563

; CURRENT APPLICATION NUMBER: US/10/066,543
; CURRENT FILING DATE: 2002-01-31

; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2964

; LENGTH: 567

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc.feature

; LOCATION: 527

; OTHER INFORMATION: n = A,T,C or G

US-10-066-543-2964

Query Match 18.5%; Score 187.8; DB 14; Length 567;
Best Local Similarity 73.3%; Pred. No. 6.6e-50;
Matches 304; Conservative 0; Mismatches 87; Indels 24; Gaps 4;

QY 600 CCCTTCCCGCTGTGCTCCCTCCCGATGGGGGTTGAGTCCATATTTAACTGGCCATC 659
DB 481 CTCCTACCTGCGATGCTCCCTCCCAATTTGGGCAATTAATCACTTTAACTGGTTGT 422
QY 660 CTCACAGTTGCTAACTAGCAAGTGTCTTTCTTTAGGACCCCTCTCTTAACGAGCAATAT 719
DB 421 CTGTTAGTCGTAACCTAGTAAGTGTCTTTCTTTATAGAACCCCTCTGACTGAGCAATAT 362
QY 720 GTCTGACCTGTACTATAAGATCTTTCTGATAATGCAATCGGAGATTTTTTGGTAGATAG 779
DB 361 GCCT-CCCTGTATTATAAATCTTTCTGATAATGCAATCGGAGATTTTTTGGTAGATAG 303
QY 780 TAGAAGTGGTTCCTGTTTTCACCTTCTTACTCAGCTGACTAGTGTCTTCCCTTCGTTT 839
DB 302 TAAAGTGTCTTCCATGTTAC-----TTTATTACAGAGCTAATAAGTGTCTTCCCTTAGTT 248
QY 840 TCTAGTAAGTGGGTGTAGAAATCACTGTCTGGGGTTTACAGTTTTTAACTATTTTGA 899
DB 247 TCTAGTAACCTAGGTGTAAATAATCATGTGTGACGCTTTATAGTTTTTAAATAATTTTGA 188
QY 900 TA-----TTCTGAACATCACTGTCTTCCAGAGTACCAACACTGTCA 942
DB 187 TAATCTTAAATATGAACCTCTTTAACTATCACTGTCTTCCAGATTTACGACACTGTCA 128
QY 943 TGTGATTGATGGCGCCCTCTAGACCTCACTACCGGAGACACATGCTTCCGGTA 997
DB 127 CTTGACCAATACCTG-ACCCTCTTTACCTCGCCCGGGACACACGCTCTCTGTA 74

RESULT 13

US-09-871-161-483/c
; Sequence 483, Application US/09871161
; Publication No. US20030097666A1
; GENERAL INFORMATION:

; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; PRODUCTS: 11

FILE REFERENCE: CDNA-260XX
CURRENT APPLICATION NUMBER: US/09/871,161
CURRENT FILING DATE: 2001-05-31
PRIOR APPLICATION NUMBER: 09/328,111
PRIOR FILING DATE: 1999-06-08
PRIOR APPLICATION NUMBER: 60/117,393
PRIOR FILING DATE: 1999-01-27
PRIOR APPLICATION NUMBER: 60/098,639
PRIOR FILING DATE: 1998-08-31
NUMBER OF SEQ ID NOS: 544
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 483
LENGTH: 605
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(605)
OTHER INFORMATION: n = A,T,C or G
US-09-871-161-483

Query Match 18.4%; Score 187.2; DB 11; Length 605;
Best Local Similarity 65.5%; Pred. No. 1.1e-49;
Matches 348; Conservative 0; Mismatches 158; Indels 25; Gaps 5;
QY 485 GTACATGGAGTAATCAACTGAATAAAGTGTCAAGGTAAGCTTTTAAACGGTTAATT 544
Db 597 GGCAANGAANNAAGTCCGGGTAAACCTTTAAANGGTTAATTTTGTCAAATNCAGTA 538
QY 545 TCTGTCAACAGTAG-ATGCAAAATGCGCGATCTTATCAGTGTCTCTTGGAGCCGCCCT 603
Db 537 GATAAANAANAAGTTTGNATAAACAGNAATGTTTCTTANGCTTTTCNTTNTCTT 478
QY 604 TCCCTCTGCTGCTCCCTCCAGATGGGGGCTTGAGTCCATATTAAACTGCCCATCTCA 663
Db 477 AACACCTGCCATGCTCCCAAAATGGGGCATTTAAATCATCTTTAAACGGTTGTTCTGT 418
QY 664 CAGTTGCTAACTTAGCAAGTCTTTCTTTAGGACCCCTTCTTAACGACCAATATGCT 723
Db 417 TAGTCGCTAACTTAGTAAGTGTCTTTCTTATAGAACCCCTTCTGACTGAGCAATATGCT 358
QY 724 GACCTGCTATTAAGATCTTTCTGATAATGCAATCGGAGATTTTGGTAGATAGTAGA 783
Db 357 -CCTTGATTAATAAATCTTCTGATAATGCAATCGAAGTCTTTTGTGCGANTAGTAA 299
QY 784 AGTGGCTTCTGTTTCACTTCTTCTTACGCTGACTAGTGTCCCTTCCCTTCTGTTCTA 843
Db 298 AGTGGCTTCTGTTTCACTTCTTCTTACGCTGACTAGTGTCCCTTCCCTTCTGTTCTA 244
QY 844 GTAACCTGGGTGTAGAATCACTGCTGCGCTTTTACAGTTTCTTAAACTATTTAGATA-- 901
Db 243 GTAACCTAGTGTAAAATCATGTTGCGAGCTTTATAGTTTTTAAATATTTTAGATAAT 184
QY 902 -----TTCTGAACATCATCTGTCTTCCAGAGTACCAACACTGTCATGTG 946
Db 183 TCTTAAACTATGAACCTTCTTAACTACCTGCTTGCAGATTAACGACACTGTCATGTG 124
QY 947 ATTGATGCGCCCTCTAGACCTCACCACGCGGACACATGCTTCGGGTA 997
Db 123 ACCAATAGT-ACCCTCTTTACTCGCCACGCGGACACACGCTCTCTGTA 74

RESULT 14
US-10-066-543-3022/c
Sequence 3022, Application US/10066543
Publication No. US20030087818A1
GENERAL INFORMATION:
APPLICANT: Jiang, Yuqiu
APPLICANT: Pyle, Ruth A.
APPLICANT: Xu, Jiangchun
APPLICANT: Indrias, Carol Yoseph
APPLICANT: Lodes, Michael J.
APPLICANT: Secrist, Heather

APPLICANT: Carter, Darrick
APPLICANT: Fanger, Gary R.
APPLICANT: Smith, Carole L.
APPLICANT: Durham, Margarita
APPLICANT: Stolk, John A.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
FILE REFERENCE: 210121.563
CURRENT APPLICATION NUMBER: US/10/066,543
CURRENT FILING DATE: 2002-01-31
NUMBER OF SEQ ID NOS: 3417
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3022
LENGTH: 477
TYPE: DNA
ORGANISM: Homo sapiens
US-10-066-543-3022

Query Match 18.2%; Score 185.4; DB 14; Length 477;
Best Local Similarity 73.2%; Pred. No. 3.5e-49;
Matches 301; Conservative 0; Mismatches 86; Indels 24; Gaps 4;
QY 604 TCCCTCTGCTGCTCCCTCCAGATGGGGCTTGAGTCCATATTTAAACTGCCCATCTCA 663
Db 477 TACACCTGCCATGCTCCCAAAATGGGCATTTAAATCATCTTTAAACTGGTGTCTGT 418
QY 664 CAGTTGCTAACTTAGCAAGTCTTTCTTTAGGACCCCTTCTTAACGACCAATATGCT 723
Db 417 TAGTCGCTAACTTAGTAAGTGTCTTTCTTATAGAACCCCTTCTGACTGAGCAATATGCT 358
QY 724 GACCTGCTATTAAGATCTTTCTGATAATGCAATCGGAGATTTTGGTAGATAGTAGA 783
Db 357 -CCTTGATTAATAAATCTTCTGATAATGCAATCGAAGTCTTTTGTGCGANTAGTAA 299
QY 784 AGTGGCTTCTGTTTCACTTCTTCTTACGCTGACTAGTGTCCCTTCCCTTCTGTTCTA 843
Db 298 AGTGGCTTCTGTTTCACTTCTTCTTACGCTGACTAGTGTCCCTTCCCTTCTGTTCTA 244
QY 844 GTAACCTGGGTGTAGAATCACTGCTGCGCTTTTACAGTTTCTTAAACTATTTAGATA-- 901
Db 243 GTAACCTAGTGTAAAATCATGTTGCGAGCTTTATAGTTTTTAAATATTTTAGATAAT 184
QY 902 -----TTCTGAACATCATCTGTCTTCCAGAGTACCAACACTGTCATGTG 946
Db 183 TCTTAAACTATGAACCTTCTTAACTACCTGCTTGCAGATTAACGACACTGTCATGTG 124
QY 947 ATTGATGCGCCCTCTAGACCTCACCACGCGGACACATGCTTCGGGTA 997
Db 123 ACCAATAGT-ACCCTCTTTACTCGCCACGCGGACACACGCTCTCTGTA 74

RESULT 15
US-09-960-352-7612
Sequence 7612, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengbing
APPLICANT: Byatt, John C.
APPLICANT: Mathialagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 7612
LENGTH: 432
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 33-LIB188-020-Q1-E1-A2
US-09-960-352-7612

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 23, 2003, 09:17:01 ; Search time 99.1939 Seconds
(without alignments)
9470.385 Million cell updates/sec

```

Title: US-09-717-321A-17
Perfect score: 348
Sequence: 1 tgaaacatcacgtcttggc.....ctttgaacttgaaaaaaa 348

```

Scoring table: IDENTITY_NUC
Gapop 10.0 . Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

```

Database :
N_Genseq_19Jun03.*
1: /SIDSL/gcgdata/genseq/genseq-emb1/NA1980.DAT.*
2: /SIDSL/gcgdata/genseq/genseq-emb1/NA1981.DAT.*
3: /SIDSL/gcgdata/genseq/genseq-emb1/NA1982.DAT.*
4: /SIDSL/gcgdata/genseq/genseq-emb1/NA1983.DAT.*
5: /SIDSL/gcgdata/genseq/genseq-emb1/NA1984.DAT.*
6: /SIDSL/gcgdata/genseq/genseq-emb1/NA1985.DAT.*
7: /SIDSL/gcgdata/genseq/genseq-emb1/NA1986.DAT.*
8: /SIDSL/gcgdata/genseq/genseq-emb1/NA1987.DAT.*
9: /SIDSL/gcgdata/genseq/genseq-emb1/NA1988.DAT.*
10: /SIDSL/gcgdata/genseq/genseq-emb1/NA1989.DAT.*
11: /SIDSL/gcgdata/genseq/genseq-emb1/NA1990.DAT.*
12: /SIDSL/gcgdata/genseq/genseq-emb1/NA1991.DAT.*
13: /SIDSL/gcgdata/genseq/genseq-emb1/NA1992.DAT.*
14: /SIDSL/gcgdata/genseq/genseq-emb1/NA1993.DAT.*
15: /SIDSL/gcgdata/genseq/genseq-emb1/NA1994.DAT.*
16: /SIDSL/gcgdata/genseq/genseq-emb1/NA1995.DAT.*
17: /SIDSL/gcgdata/genseq/genseq-emb1/NA1996.DAT.*
18: /SIDSL/gcgdata/genseq/genseq-emb1/NA1997.DAT.*
19: /SIDSL/gcgdata/genseq/genseq-emb1/NA1998.DAT.*
20: /SIDSL/gcgdata/genseq/genseq-emb1/NA1999.DAT.*
21: /SIDSL/gcgdata/genseq/genseq-emb1/NA2000.DAT.*
22: /SIDSL/gcgdata/genseq/genseq-emb1/NA2001A.DAT.*
23: /SIDSL/gcgdata/genseq/genseq-emb1/NA2001B.DAT.*
24: /SIDSL/gcgdata/genseq/genseq-emb1/NA2002.DAT.*
25: /SIDSL/gcgdata/genseq/genseq-emb1/NA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES							
Result No.	Score	Query			DB	ID	Description
		Match	Length	%			
1	348	100.0	348	22	AAH223399	Human rac1 gene re	
C 2	348	100.0	1266	22	AAH22336	Human rac1 contig	
3	209.4	60.2	2051	21	AAH182336	Lung cancer associ	
4	203.6	58.5	28567	25	ABY17030	Human M21 gene Ra	
5	187.8	54.0	352	22	AAH22400	Human rac1 gene re	
6	187.4	53.9	1232	24	ABX83623	Human cDNA differe	
7	187.4	53.9	1232	24	ABN951134	Gene #1632 used to	
8	187.4	53.9	1318	23	AA565569	DNA encoding novel	

ALIGNMENTS

RESULT 1

AAH22399
ID AAH22399 standard; DNA; 348 BP.

AAH22399;

DT 22-AUG-2001 (first entry)

DE Human rac1 gene related nucleotide sequence #3.

XX Identification; toxic; hepatotoxic; differential gene expression;
KW NSAID; non-steroidal antiinflammatory drug; ds.
KW

OS Homo sapiens.

AA WO200138579-A2

XX 31-MAY-2001 PD

21-NOV-2000: 2000WO-IIS32049

22-NOV-1999: 99IIS-0166923

PR 18-FEB-2000; 2000US-0183531.
PR 20-NOV-2000; 2000US-0717321

AA
PA (CURA-) CURAGEN CORP.

XX Gould-Rothberg BE. D

XX
DR WPI: 2001-355948/37.

WY

PT Screening hepatotoxic agent comprises contacting test cell population
PT expressing RISKMARKER or INJURYMARKER with agent, comparing expression
PT with reference population and identifying difference in expression
PT levels
XX
XX Disclosure; Page 8-9; 76pp; English.
XX
XX The present invention describes a method of screening a test agent for
CC hepatotoxicity. The method comprises: (a) providing a test cell
CC population comprising a cell capable of expressing one or more nucleic
CC acid sequences selected from the group consisting of RISKMARKER 1-8
CC and INJURYMARKER 1-10; (b) contacting the test cell population with a
CC test agent; (c) measuring expression of one or more of the nucleic
CC acid sequences in the test cell population; (d) comparing the
CC expression of the nucleic acid sequence in the test cell population to
CC the expression of the nucleic acid sequence in an reference cell
CC population comprising at least one cell whose exposure status to a
CC hepatotoxic agent in known; and (e) identifying a difference in
CC expression levels of the RISKMARKER or INJURYMARKER sequences, if
CC present, in the test cell population and reference cell population.
CC The method is useful for identifying a hepatotoxic agent. The present
CC sequence is given in the exemplification of the present invention.
XX
SQ Sequence 348 BP; 103 A; 76 C; 67 G; 102 T; 0 other;

Query Match 100.0%; Score 348; DB 22; Length 348;
Best Local Similarity 100.0%; Pred. No. 2.5e-77;
Matches 348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAACATCACTGCTTGGCAGAGTACCAACACTGTCATGTGATGCGCCCTCT 60
Db 1 TGAACATCACTGCTTGGCAGAGTACCAACACTGTCATGTGATGCGCCCTCT 60

QY 61 AGACCTCACCCACGCGGACACATGCTTCGGTACCTTTGGTCTGTGAGTTCTGTCAAG 120
Db 61 AGACCTCACCCACGCGGACACATGCTTCGGTACCTTTGGTCTGTGAGTTCTGTCAAG 120

QY 121 CGTAGTGTAAACGCGGTTCTGTACAACTTACTCTGCGAAGACACAGTTGGGCC 180
Db 121 CGTAGTGTAAACGCGGTTCTGTACAACTTACTCTGCGAAGACACAGTTGGGCC 180

QY 181 TTTCGACCACTAGAACAAACTTTTTCATTTGACAGTTCAGAAATGTGAGTGTTTTA 240
Db 181 TTTCGACCACTAGAACAAACTTTTTCATTTGACAGTTCAGAAATGTGAGTGTTTTA 240

QY 241 CATTGATCTTTTCTAATGCAGTTAGCAGTATGTTTTCATGATGATGATTAATAATCCT 300
Db 241 CATTGATCTTTTCTAATGCAGTTAGCAGTATGTTTTCATGATGATGATTAATAATCCT 300

QY 301 TGAATCATATAAAAAAATAAATGCTTTTGGAACTTGAAAAAAA 348
Db 301 TGAATCATATAAAAAAATAAATGCTTTTGGAACTTGAAAAAAA 348

RESULT 2
AAH22396/c
ID AAH22396 standard; DNA; 1266 BP.
XX
XX
AC AAH22396;
XX
XX
DT 22-AUG-2001 (first entry)
XX
XX Human rac1 contig SEQ ID NO:2.
DE
XX
XX Identification; toxic; hepatotoxic; differential gene expression;
KW NSAID; non-steroidal antiinflammatory drug; ds.
XX
XX Homo sapiens.
OS
XX
XX WO200138579-A2.
PN
XX
XX 31-MAY-2001.
PD
XX

PF 21-NOV-2000; 2000WO-US32049.
XX
XX 22-NOV-1999; 99US-0166923.
PR 18-FEB-2000; 2000US-0183531.
PR 20-NOV-2000; 2000US-0717321.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Gould-Rothberg BE, Dipippo VA, Ramesh TM, Gerwein RW;
PI WPI; 2001-355948/37.
XX
XX Screening hepatotoxic agent comprises contacting test cell population
PT expressing RISKMARKER or INJURYMARKER with agent, comparing expression
PT with reference population and identifying difference in expression
PT levels
XX
XX Disclosure; Page 7; 76pp; English.
PS
XX The present invention describes a method of screening a test agent for
CC hepatotoxicity. The method comprises: (a) providing a test cell
CC population comprising a cell capable of expressing one or more nucleic
CC acid sequences selected from the group consisting of RISKMARKER 1-8
CC and INJURYMARKER 1-10; (b) contacting the test cell population with a
CC test agent; (c) measuring expression of one or more of the nucleic
CC acid sequences in the test cell population; (d) comparing the
CC expression of the nucleic acid sequence in the test cell population to
CC the expression of the nucleic acid sequence in an reference cell
CC population comprising at least one cell whose exposure status to a
CC hepatotoxic agent in known; and (e) identifying a difference in
CC expression levels of the RISKMARKER or INJURYMARKER sequences, if
CC present, in the test cell population and reference cell population.
CC The method is useful for identifying a hepatotoxic agent. The present
CC sequence is given in the exemplification of the present invention.
XX
SQ Sequence 1266 BP; 385 A; 258 C; 285 G; 338 T; 0 other;

Query Match 100.0%; Score 348; DB 22; Length 1266;
Best Local Similarity 100.0%; Pred. No. 3.5e-77;
Matches 348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAACATCACTGCTTGGCAGAGTACCAACACTGTCATGTGATGCGCCCTCT 60
Db 357 TGAACATCACTGCTTGGCAGAGTACCAACACTGTCATGTGATGCGCCCTCT 298

QY 61 AGACCTCACCCACGCGGACACATGCTTCGGTACCTTTGGTCTGTGAGTTCTGTCAAG 120
Db 297 AGACCTCACCCACGCGGACACATGCTTCGGTACCTTTGGTCTGTGAGTTCTGTCAAG 238

QY 121 CGTAGTGTAAACGCGGTTCTGTACAACTTACTCTGCGAAGACACAGTTGGGCC 180
Db 237 CGTAGTGTAAACGCGGTTCTGTACAACTTACTCTGCGAAGACACAGTTGGGCC 178

QY 181 TTTCGACCACTAGAACAAACTTTTTCATTTGACAGTTCAGAAATGTGAGTGTTTTA 240
Db 177 TTTCGACCACTAGAACAAACTTTTTCATTTGACAGTTCAGAAATGTGAGTGTTTTA 118

QY 241 CATTGATCTTTTCTAATGCAGTTAGCAGTATGTTTTCATGATGATGATTAATAATCCT 300
Db 117 CATTGATCTTTTCTAATGCAGTTAGCAGTATGTTTTCATGATGATGATTAATAATCCT 58

QY 301 TGAATCATATAAAAAAATAAATGCTTTTGGAACTTGAAAAAAA 348
Db 57 TGAATCATATAAAAAAATAAATGCTTTTGGAACTTGAAAAAAA 10

RESULT 3
AAF18236
ID AAF18236 standard; DNA; 2051 BP.
XX
XX AAF18236;
XX
XX 14-MAR-2001 (first entry)
DT

XX DE Lung cancer associated polynucleotide sequence SEQ ID 255.
XX DE Human; lung cancer associated protein; neuroprotective; cytostatic;
XX KW cardioactive; immunomodulatory; muscular active; vulnerary;
XX KW gastrointestinal; nephrotropic; antiinfective; gynecological;
XX KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
XX KW proliferative disorder; wound healing; infectious disease; ds.
XX OS Homo sapiens.
XX PN WO200055190-A2.
XX PD 21-SEP-2000.
XX PF 08-MAR-2000; 2000WO-US05919.
XX PR 12-MAR-1999; 99US-0124270.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PA (ROSE/) ROSEN C A.
XX PI Ruben SM;
XX DR WPI; 2000-587514/55.
XX DR P-PSDB; AAB58360.
XX PT Lung cancer associated gene sequences, referred to as lung cancer
XX PT antigens, useful for treatment, prevention, and diagnosis of disorders
XX PT such as lung cancer -
XX PS Claim 1; Page 716-717; 1425pp; English.
XX CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
XX CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
XX CC associated proteins and polynucleotide sequences, their agonists, and
XX CC antagonists may have neuroprotective; cytostatic; cardioactive;
XX CC immunomodulatory; muscular active general; vulnerary; gastrointestinal
XX CC general; nephrotropic; antiinfective; gynecological; or antibacterial
XX CC activity. The invention also includes antibodies specific for the
XX CC protein or polynucleotide sequences. The lung cancer associated
XX CC polynucleotide sequences may be used for detection of lung cancer,
XX CC chromosome identification, as chromosome markers, and for numerous other
XX CC diagnostic or research purposes. The proteins may be used to treat
XX CC disorders such as neural, immune, muscular, reproductive,
XX CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
XX CC disorders. The proteins may also be used in the treatment of wounds and
XX CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
XX CC peptide AAB58549 are used in the course of the invention for the
XX CC identification and characterisation of the polynucleotide and protein
XX CC sequences.
XX SQ Sequence 2051 BP; 570 A; 430 C; 433 G; 612 T; 6 other;
Query Match 60.2%; Score 209.4; DB 21; Length 2051;
Best Local Similarity 82.6%; Pred. No. 1 4e-42;
Matches 285; Conservative 0; Mismatches 36; Indels 24; Gaps 3;
QY 4 AACATCACTGCTTGGCAGAGTACCAACACTGTCATGTGATGATGCGGCCCTCTAGA 63
Db 1666 AACATCACTGCTTGGCAGATTACCGACACTGTCACCTTGACCAATCTGA-CCCTCTTTA 1724
QY 64 CCTCACCCACGGGACACATGCTCCGGTA-----CCTTGGGT 102
Db 1725 CCTCGCCACCGGACACACGCTCTGTGTAGTCGCTTTGTGCTATGATGTTCCCTTGGGT 1784
QY 103 CTGTGAGGTTCTGTCAA--CGGTAGTGTAAACGCGTTCCTGTACACCTAACTCACTGG 160
Db 1785 CTGTGAGGTTCTGTAACTGTTGTAGTGTGACGATGTTCTGTACACTTAACTCACTGG 1844
QY 161 CAGAACACAGTGTGGGCGCTTTCGACCACTAGAACAAAATTTTTCAAATTGACAGTTGC 220
Db 1845 CGAAGATACAGCTGGGACCTTCAGCCACTACACAGAAATTTTAAATTGACAGTTGC 1904

QY 221 AGAATTGTGGAGTGTCTTTTACATTGAICTTTTGCCTAATGCAGTTAGCAGTATGTTTGA 280
Db 1905 AGAATTGTGGAGTGTCTTTTACATTGATCTTTTGTCTAATGCAATTAGCATTATGTTTGA 1964
QY 281 TGTATGACTTAAATAATCCCTTGAATCATATAAAAAAAAAAAAAA 325
Db 1965 TGTATGACTTAAATAATCCCTTGAATCATATAAAAAAAAAAAAAA 2009

RESULT 4

ABT17030
ID ABT17030 standard; DNA; 28567 BP.
XX AC ABT17030;
XX DT 03-APR-2003 (first entry)
XX DE Human MP21 gene Racl SEQ ID No 4.
XX KW Cytostatic; p21 pathway modulating agent; cancer; angiogenic; apoptotic;
XX KW cell proliferation disorder; MP21; gene; ds.
XX OS Homo sapiens.
XX PN WO2003006990-A1.
XX PD 23-JAN-2003.
XX PF 10-JUL-2002; 2002WO-US21549.
XX PR 12-JUL-2001; 2001US-305017P.
XX PR 10-OCT-2001; 2001US-328491P.
XX PR 15-FEB-2002; 2002US-357452P.
XX PA (EXEL-) EXELIXIS INC.
XX PI Friedman L, Plowman GD, Belvin M, Li D, Funke RP;
XX DR WPI; 2003-221779/21.
XX DR P-PSDB; ABJ19756.
XX PT Identifying candidate p21 pathway modulator, by contacting an assay
XX PT system having modifiers of p21 polypeptide or gene with a test agent to
XX PT provide a reference activity in system and detecting test agent-biased
XX PT activity -
XX PS Examples; Page 56-72; 199pp; English.
XX CC The invention relates to a novel method for identifying a candidate p21
XX CC pathway modulating agent. The novel method comprises contacting an assay
XX CC system, comprising a purified MP21 polypeptide (modifier of p21) or
XX CC nucleic acid, with a test agent under conditions, so that but for the
XX CC presence of a test agent, the assay system provides a reference activity
XX CC and detection of test agent-biased activity of the assay system. The
XX CC novel method of the invention is useful for identifying a candidate p21
XX CC pathway modulating agent. The invention also includes a method for
XX CC modulating the p21 pathway of a cell, and a method for diagnosing a
XX CC disease e.g. cancer in a patient. The identified modulators are useful in
XX CC diagnosis, therapy and pharmaceutical development. The modulators are
XX CC useful in a variety of diagnostic and therapeutic applications including
XX CC angiogenic, apoptotic and cell proliferation disorders. This
XX CC polynucleotide sequence represents a gene encoding an MP21 protein of the
XX CC invention.
XX SQ Sequence 28567 BP; 6762 A; 6358 C; 6796 G; 8651 T; 0 other;
Query Match 58.5%; Score 203.6; DB 25; Length 28567;
Best Local Similarity 78.8%; Pred. No. 7.5e-41;
Matches 290; Conservative 0; Mismatches 54; Indels 24; Gaps 3;
QY 4 AACATCACTGCTTGGCAGAGTACCAACACTGTCATGTGATGATGCGGCCCTCTAGA 63

Db 27782 AACATCACTGCTTCCGAGATTACCGACACTGCTCATTGACCAATACTGA-CCCTCTTTA 27840
QY 64 COTACCCACGCGGACACATGTTCCGTA-----CCTTGGGT 102
Db 27841 COTGCGCCACGCGGACACAGCCTCCTGTAGTCGCTTTGCTTATTGATGTTCTTGGGT 27900
QY 103 CTGTGAGTCTGTCAA--GCCTAGTGTAAAGCGCTTCTGTACAACTTAACCTCACTGG 160
Db 27901 CTGTGAGGTCTGTAACTGTCTAGTGTGACGATGTTCTGTACAACTTAACCTCACTGG 27960
QY 161 CAAGAACACAGTGTGGGCTTTGACCACTAGAACAAACTTTTCAATTGACAGTTGC 220
Db 27961 CGAGATACAGCTGGGACCTTCAGCCACTACACAGAAATTTTAAATTGACAGTTGC 28020
QY 221 AGAATTGTGAGTGTGTTTACATTTGATGCTTTTGTCTTAATGAGTGTAGCAGTATGTTTGA 280
Db 28021 AGAATTGTGAGTGTGTTTACATTTGATGCTTTTGTCTTAATGCAATTAGCATTTATGTTTGA 28080
QY 281 TGTATGACTTAATAATCCTTGAATCATATAAAAAAATAAATGCTTTTGGACTTG 340
Db 28081 TGTATGACTTAATAAATCCTTGAATCATACGACTGCTGTAATACTGGTGTGTTTGGAGCTTG 28140
QY 341 AAAAAAA 348
Db 28141 ATGAACAA 28148
RESULT 5
ID AAH22400 standard; DNA; 352 BP.
XX AC AAH22400;
XX AC AAH22400;
DT 22-AUG-2001 (first entry)
XX Human rac1 gene related nucleotide sequence #4.
XX Identification; toxic; hepatotoxic; differential gene expression;
KW NSAID; non-steroidal antiinflammatory drug; ds.
XX Homo sapiens.
XX WO200138579-A2.
XX 31-MAY-2001.
XX 21-NOV-2000; 2000WO-US32049.
XX 22-NOV-1999; 99US-0166923.
PR 18-FEB-2000; 2000US-0183531.
PR 20-NOV-2000; 2000US-0717321.
XX (CURA-) CURAGEN CORP.
PA Gould-Rothberg BE, Dipippo VA, Ramseh TM, Gerwein RW;
PI WPI; 2001-355948/37.
XX Screening hepatotoxic agent comprises contacting test cell population
PT expressing RISKMARKER or INJURYMARKER with agent, comparing expression
PT with reference population and identifying difference in expression
PT levels -
XX Disclosure; Page 8-9; 76pp; English.
XX The present invention describes a method of screening a test agent for
CC hepatotoxicity. The method comprises: (a) providing a test cell
CC population comprising a cell capable of expressing one or more nucleic
CC acid sequences selected from the group consisting of RISKMARKER 1-8
CC and INJURYMARKER 1-10; (b) contacting the test cell population with a
CC test agent; (c) measuring expression of one or more of the nucleic
CC acid sequences in the test cell population; (d) comparing the
CC expression of the nucleic acid sequence in the test cell population to

CC the expression of the nucleic acid sequence in an reference cell
CC population comprising at least one cell whose exposure status to a
CC hepatotoxic agent in known; and (e) identifying a difference in
CC expression levels of the RISKMARKER or INJURYMARKER sequences, if
CC present, in the test cell population and reference cell population.
CC The method is useful for identifying a hepatotoxic agent. The present
CC sequence is given in the exemplification of the present invention.
XX Sequence 352 BP; 86 A; 77 C; 72 G; 117 T; 0 other;
SQ Query Match 54.0%; Score 187.8; DB 22; Length 352;
Best Local Similarity 77.6%; Pred. No. 2.1e-37;
Matches 274; Conservative 0; Mismatches 57; Indels 22; Gaps 3;
QY 17 TCCAGAGTACCAACACTGTCATGTGATGATGCGCCGCCCTCTAGACCTCACCCACGG 76
Db 1 TCCAGAGTACCGACACTGTCACCTTGACCAATACG-ACCCTCTTTACCTCGCCACGG 59
QY 77 GACATGCTTCG-----GTACCTTTGGGTCTGTGAGTTCTGTC 117
Db 60 GACACCGCTCTGCTGCTCGCTTGGCTATTGATGTTCCITTTGGTCTGTGAGTTCTGTA 119
QY 118 AA--GCGTAGTCTAAACCGCTTCTGTACACCTTAACCTCACTGGCAAGACACAGTGT 175
Db 120 AACTGTGCTAGTCTGACGATGTTCTGTACAACTTAACCTCACTGGCGAGAAACAGCGTG 179
QY 176 GGGCTTTTCGACACTAGAACAACTTTTCAATTGACAGTTGCAAGATTGGAGTGT 235
Db 180 GGACCCCTTCAGCCACTACACAGAAATTTTAAATTGACAGTTGCAAGATTGGAGTGT 239
QY 236 TTTTACATTGATCTTTTGTCTTAATGACAGTTAGCAGTATGTTTGCATGTATGACTTAATA 295
Db 240 TTTTACATTGATCTTTTGTCTTAATGCAATTAGCATTATGTTTGCATGTATGACTTAATA 299
QY 296 ATCTTTGAATCATAAAAAATAAATAAATAAATGCTTTTGGAACTTGAATAA 348
Db 300 ATCTTTGAATCATACGACTGGTAAATAGTGGTGTGTTTGGAACTTGAATAA 352
RESULT 6
ID ABK83623
XX ABK83623 standard; cDNA; 1232 BP.
XX AC ABK83623;
XX AC ABK83623;
DT 14-AUG-2002 (first entry)
XX Human cDNA differentially expressed in granulocytic cells #194.
XX Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW viral infection; parasitic infection; protozoal infection;
KW fungal infection; sterile inflammatory disease; psoriasis;
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW adult respiratory distress syndrome; inflammatory bowel disease;
KW Crohn's disease; ulcerative colitis; periodontal disease;
KW granulocyte activation; chronic inflammation; allergy.
XX Homo sapiens.
OS WO200228999-A2.
XX WO200228999-A2.
XX 11-APR-2002.
XX 03-OCT-2001; 2001WO-US30821.
XX 03-OCT-2000; 2000US-237189P.
XX (GENE-) GENE LOGIC INC.
XX Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX WPI; 2002-435328/46.
DR

	XX	Detecting granulocyte activation by detecting differential expression
	PT	of genes associated with granulocyte activation, which serves as
	PT	diagnostic markers that is useful for monitoring disease states and
	PT	drug toxicity -
	XX	
	PS	Claim 1; SEQ ID No 194; 114bp; English.
	XX	
	CC	The invention relates to detecting (M1) granulocyte (GC) activation
	CC	(GCA), by detecting the level of expression of gene(s) (Gs) identified by
	CC	DNA chip analysis as given in the specification, and comparing
	CC	the expression level to an expression level in an unactivated
	CC	GC, where differential expression of Gs is indicative of GCA.
	CC	Also included are modulation of at least one gene in Gs; (2) screening (M3)
	CC	that alters the expression of GCA by contacting GC with an agent
	CC	for an agent capable of modulating GCA or an inflammation (especially
	CC	chronic) in a tissue, an allergic response in a subject, exposure of a
	CC	subject to a pathogen or sterile inflammatory disease using the
	CC	gene expression profile; (3) detecting (M4) an inflammation (especially
	CC	chronic) in a tissue, an allergic response in a subject, exposure of a
	CC	subject to a pathogen or sterile inflammatory disease, by detecting the
	CC	level of expression in a sample of the tissue of gene(s) from Gs, where
	CC	the level of expression of the gene is indicative of inflammation;
	CC	(4) treating (M5) an inflammation (especially chronic) or in a tissue,
	CC	an allergic response in a subject, exposure of a subject to a pathogen
	CC	or sterile inflammatory disease, by contacting a tissue having
	CC	inflammation with an agent that modulates the expression of gene(s)
	CC	from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
	CC	GCA preferably in an inflammation in a tissue; M3 is useful for
	CC	detecting an inflammation (especially chronic) in a tissue, an allergic
	CC	response in a subject, exposure of a subject to a pathogen or sterile
	CC	inflammatory disease (e.g. psoriasis, rheumatoid arthritis, renal
	CC	glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
	CC	refusion injury, ARRS, adult respiratory distress syndrome,
	CC	inflammatory bowel disease, Crohn's disease, ulcerative colitis,
	CC	periodontal disease; also bacterial infection, viral infection,
	CC	parasitic infection, protozoal infection, fungal infection and M5 is
	CC	useful for treating one of the above conditions. The present
	CC	sequence represents a gene differentially expressed in granulocytes.
	CC	Note: The sequence data for this patent did not form part
	CC	of the printed specification, but was obtained in electronic
	CC	format directly from WIPO at
	CC	ftp.wipo.int/pub/published_pct_sequences.
	XX	
	SQ	Sequence 1232 BP; 329 A; 240 C; 251 G; 411 T; 1 other;
		Query Match 53.9%; Score 187.4; DB 24; Length 1232;
		Best Local Similarity 80.9%; Pred. No. 3.6e-37;
		Matches 263; Conservative 1; Mismatches 37; Indels 24; Gaps 3;
	QY	4 AACATCACTGTCTTGCCAGAGTACCACACTGTCATGTGAATGATGGCGGCCCTCTAGA 63
	Db	905 AACATCACTGTCTTGCCAGATTACCGACACTGTCACCTTGACCAATACTGA-CCTCTTTA 963
	QY	64 CCTCACCCAGCGGACACATGCTTCCGGTA-----CCTTTGGGT 102
	Db	964 CCTCGGCCACGGGACACACGCTCTCTGTAGTCGCTTGGCTATTGATGTTCCITTGGT 1023
	QY	103 CTGTGAGGTTCTGTCAA--GGGTAGTGTCTAAAGCCGCTTCGTACAACTCACTCTG 160
	Db	1024 CTGTGAGGTTCTGTAAACTGTGTAGTGCTGACGATGCTCTGTACAACTTAATCACTG 1083
	QY	161 CAAGAACACAGTGTGGGCCCTTCGACCACCTAGNACAACATTTTTTCAATGACAGTTC 220
	Db	1084 CGAATATACAGCTGGGACCCCTTCAGCCACTCAACAGAAATTTTTAAATTGACAGTTC 1143
	QY	221 AGAATTGTGGAGTGTTTTACATTGATCTTTTGTCAATGCAGTTAGCAGTATCTTTGCA 280
	Db	1144 AGAATTGTGGAGTGTTTTACATTGATCTTTTGTCAATGCAGTATCTTTGCA 1203
	QY	281 TGATGACTTAAATAATCTCTTGAAT 305

QY 161 CAAGACACAGTGTGGCCCTTTCAGACACTAGAACAACTTTTCAATTGACAGTTGC 220
 Db 1084 CGAAGATACAGCGTGGGACCCCTTCAAGCCACTACACAGAAATTTTAAATTGACAGTTGC 1143
 QY 221 AGAATTGTGGAGTGTGTACATTCATCTTTTGTCTAATGCACTTAGCAGTATGTTTTCGA 280
 Db 1144 AGAATTGTGGAGTGTGTACATTCATCTTTTGTCTAATGCACTTAGCAGTATGTTTTCGA 1203
 QY 281 TGTATGACTTAAATAATCCTTGAAT 305
 Db 1204 TGTATGACTTAAATAATCCTTGAAT 1228

RESULT 8
 ID AAS65569 standard; cDNA; 1318 BP.
 XX AAS65569;
 AC AAS65569;
 XX 13-FEB-2002 (first entry)
 DT 13-FEB-2002 (first entry)
 DE DNA encoding novel human diagnostic protein #1373.
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX Homo sapiens.
 OS WO200175067-A2.
 XX 11-OCT-2001.
 XX 30-MAR-2001; 2001WC-US08631.
 XX 31-MAR-2000; 2000US-0540217.
 XX 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSEQ INC.
 XX Dmanac RT, Liu C, Tang YT;
 XX WPI; 2001-639362/73.
 XX P-PSDB; ABG01382.

XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX Claim 1; SEQ ID No 1373; 103pp; English.
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 1318 BP; 346 A; 268 C; 279 G; 424 T; 1 other;
 Query Match 53.9%; Score 187.4; DB 23; Length 1318;
 Best Local Similarity 80.9%; Pred. No. 3.7e-37;
 Matches 263; Conservative 1; Mismatches 37; Indels 24; Gaps 3;
 QY 4 AACATCACTGCTTCCGAGATACCAACACTGTCTCATGTGATGTGCGCCGCCCTCTAGA 63
 Db 991 AACATCACTGCTTCCGAGATACCAACACTGTCTCATGTGATGTGCGCCGCCCTCTAGA 1049
 QY 64 COTCAACCCACGCGGACACATGCTTCGGTA-----CCTTTGGGT 102
 Db 1050 COTCAACCCACGCGGACACATGCTTCGGTA-----CCTTTGGGT 1109
 QY 103 CTGTGAGGTTCTGTCAA--GCGCTAGTGTCTAAGCGGCTCTCTGACAACTAACTCACTGG 160
 Db 1110 CTGTGAGGTTCTGTAACTAGTGTCTAGTGTCTGAGATGTTCTGTACAACTTAACCTACTGG 1169
 QY 161 CAAGAACACAGTGTGGGCTTTCGACCACTAGAACAACTTTTCAATTGACAGTTGC 220
 Db 1170 CGAGAAATACAGCGTGGGACCCCTTCAGCCACTACACAGAAATTTTAAATGACAGTTGC 1229
 QY 221 AGAATTGTGGAGTGTGTACATTCATCTTTTGTCTAATGCACTTAGCAGTATGTTTTCGA 280
 Db 1230 AGAATTGTGGAGTGTGTACATTCATCTTTTGTCTAATGCACTTAGCAGTATGTTTTCGA 1289
 QY 281 TGTATGACTTAAATAATCCTTGAAT 305
 Db 1290 TGTATGACTTAAATAATCCTTGAAT 1314

RESULT 9
 ID ABX46267/c standard; cDNA; 353 BP.
 XX ABX46267;
 XX 21-FEB-2003 (first entry)
 DT 21-FEB-2003 (first entry)
 DE Bovine EST associated with lactation/muscle/fat deposition #11432.
 XX Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
 KW muscle deposition; fat deposition; genome mapping; gene identification;
 KW gene analysis; cattle breeding.
 XX Bos Taurus.
 XX US2002137139-A1.
 XX 26-SEP-2002.
 XX 24-SEP-2001; 2001US-0960352.
 XX 12-JAN-1999; 99US-115707P.
 XX 11-JAN-2000; 2000US-0480902.
 XX (BYAT/) BYATT J C.
 XX (MATH/) MATHIALAGAN N.
 XX (TAON/) TAO N.
 XX (WARR/) WARREN W C.
 XX Byatt JC, Mathialagan N, Tao N, Warren WC;
 XX WPI; 2003-110599/10.
 XX New nucleic acid associated with lactation, and muscle and fat
 PT deposition, useful for genome mapping, gene identification and
 PT analysis, cattle breeding, or for genetically improving cattle
 XX Claim 2; SEQ ID No 11432; 245pp; English.
 XX The invention relates to a purified nucleic acid molecule associated with

QY 90 GGTACCTTTGGCTGTGAGGTTCTGTCAAGCGCTAGTGTAAAGCGGTTCTGTACAACC 149
 Db |||||
 123 GGTACCTTTGGCTGTGAGGTTCTGTCAAGCGCTAGTGTAAAGCGGTTCTGTACAACC 64
 QY 150 TAACTCAGTGGCAAGACACAGTGTGGGCGCTTTCGACCACTAGAACAACTTTTTCAA 209
 Db |||||
 63 TAACTCAGTGGCAAGACACAGTGTGGGCGCTTTCGACCACTAGAACAACTTTTTCAA 4
 QY 210 TTG 212
 Db |||||
 3 TTG 1

RESULT 13
 AAL07296
 ID AAL07296 standard; DNA; 3740 BP.
 XX
 AC AAL07296;
 XX
 DT 21-NOV-2001 (first entry)
 XX
 DE Human reproductive system related antigen DNA SEQ ID NO: 9984.
 XX
 KW Human; reproductive system related antigen; reproductive system disorder;
 KW cancer; gene therapy; ds.
 XX
 OS Homo sapiens.
 XX
 PN W0200155320-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01339.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226688.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 13-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.

```
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-465570/50.
XX
XX Isolated nucleic acid molecule encoding a reproductive system antigen
PT is used in preventing, treating or ameliorating a medical condition -
XX
XX Disclosure; SEQ ID NO 984; 1297pp + Sequence Listing; English.
XX
XX The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.
XX
XX Sequence 3740 BP; 1188 A; 752 C; 687 G; 1113 T; 0 other;
SQ
Query Match 33.4%; Score 116.4; DB 22; Length 3740;
Best Local Similarity 82.0%; Pred. No. 2.4e-19;
Matches 146; Conservative 0; Mismatches 31; Indels 1; Gaps 1;
QY 147 ACCTAACTCACTGGCAGAACACAGTGTGGGCTTTTCGACCCTAGAACAAACTTTTTT 206
Db 2667 ACCTTAAGCACTGGCAGAAATATACACTGGGCTTTCACAACTAGAACAAA-TATTTT 2725
QY 207 CAATTGACAGTTGCAGAAATGTGGAGTGTGTTTACATTGATCTTTGCTAATGCAGTTAG 266
Db 2726 AATGACAGATTCAGAAATGTGGGTATTTTACATTGATCTTTGCTAATGCAGTTAG 2785
QY 267 CAGTATGTTTGCATGTATGACATTAATAATCCTTGAATCAATAAAAAAAAAAAAAA 324
Db 2786 CAATGTGTTTGCACATGGAATTAATAATCCTTGAATCAATAATAATAATAATAA 2843
RESULT 14
AAH22397
ID AAH22397 standard; DNA; 1017 BP.
XX
AC AAH22397;
XX
DT 22-AUG-2001 (first entry)
XX
XX Human rac1 gene related nucleotide sequence #1.
DE Identification; toxic; hepatotoxic; differential gene expression;
KW NSAID; non-steroidal antiinflammatory drug; ds.
XX
```

```
XX Homo sapiens.
OS WO200138579-A2.
XX
XX 31-MAY-2001.
XX
XX 21-NOV-2000; 2000WO-US32049.
XX
XX 22-NOV-1999; 9US-0166923.
PR 18-FEB-2000; 2000US-0183531.
PR 20-NOV-2000; 2000US-0717321.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Gould-Rothberg BE, Dipippo VA, Ramseh TM, Gerwein RW;
XX
XX WPI; 2001-355948/37.
XX
XX Screening hepatotoxic agent comprises contacting test cell population
PT expressing RISKMARKER or INJURYMARKER with agent, comparing expression
PT with reference population and identifying difference in expression
PT levels -
XX
XX Disclosure; Page 7-8; 76pp; English.
XX
XX The present invention describes a method of screening a test agent for
CC hepatotoxicity. The method comprises: (a) providing a test cell
CC population comprising a cell capable of expressing one or more nucleic
CC acid sequences selected from the group consisting of RISKMARKER 1-8
CC and INJURYMARKER 1-10; (b) contacting the test cell population with a
CC test agent; (c) measuring expression of one or more of the nucleic
CC acid sequences in the test cell population; (d) comparing the
CC expression of the nucleic acid sequence in the test cell population to
CC the expression of the nucleic acid sequence in a reference cell
CC population comprising at least one cell whose exposure status to a
CC hepatotoxic agent is known; and (e) identifying a difference in
CC expression levels of the RISKMARKER or INJURYMARKER sequences, if
CC present, in the test cell population and reference cell population.
CC The method is useful for identifying a hepatotoxic agent. The present
CC sequence is given in the exemplification of the present invention.
XX
XX Sequence 1017 BP; 245 A; 245 C; 216 G; 311 T; 0 other;
SQ
Query Match 32.5%; Score 113; DB 22; Length 1017;
Best Local Similarity 100.0%; Pred. No. 1.2e-18;
Matches 113; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGAACATCACTGTCTTGGCCAGAGTACCAACACTGTCATGTGATGCGGCCCTCT 60
Db 905 TGAACATCACTGTCTTGGCCAGAGTACCAACACTGTCATGTGATGCGGCCCTCT 964
QY 61 AGACCTCACCCAGCGGACACATGCTCCGGTACCTTGGTCTGTGAGTTC 113
Db 965 AGACCTCACCCAGCGGACACATGCTCCGGTACCTTGGTCTGTGAGTTC 1017
RESULT 15
ABX39192/c
ID ABX39192 standard; cDNA; 201 BP.
XX
XX ABX39192;
XX
XX 20-FEB-2003 (first entry)
XX
XX Bovine EST associated with lactation/muscle/fat deposition #4357.
XX
XX Bovine; ss; EST; expressed sequence tag; lactation; LMFP;
KW muscle deposition; fat deposition; genome mapping; gene identification;
KW gene analysis; cattle breeding.
XX
XX Bos Taurus.
OS
XX
```


PN US2002137139-A1.
XX
PD 26-SEP-2002.
XX
XX 24-SEP-2001; 2001US-0960352.
PF
XX 12-JAN-1999; 99US-115707P.
PR
XX 11-JAN-2000; 2000US-0480902.
XX
XX (BYATT/) BYATT J C.
PA (MATH/) MATHIALAGAN N.
PA (TAON/) TAO N.
PA (WARR/) WARREN W C.
XX
PI Byatt JC, Mathialagan N, Tao N, Warren WC;
XX
XX MPI; 2003-110599/10.
XX
XX New nucleic acid associated with lactation, and muscle and fat
PT deposition, useful for genome mapping, gene identification and
PT analysis, cattle breeding, or for genetically improving cattle
XX
PS Claim 2; SEQ ID No 4357; 245pp; English.
XX
XX The invention relates to a purified nucleic acid molecule associated with
CC lactation or muscle and fat deposition (designated LMFD), derived
CC from cattle, and the LMFD nucleic acid can specifically hybridise to a
CC second nucleic acid molecule comprising any of 15112 nucleotide
CC sequences, appearing as ABX34836-ABX49947, or complements of them.
CC Also included are; (1) a transformed cell having a nucleic acid
CC comprising an LMFD nucleic acid linked to a promoter and a 3' non-
CC translated sequence that functions in the cell to cause termination of
CC transcription and addition of polyadenylated ribonucleotides to a 3' end
CC of the mRNA molecule; and (2) determining a level or pattern of a
CC molecule in a bovine cell or tissue comprising: (a) incubating a marker
CC nucleic acid (comprising any of the 15112 nucleic acid sequences or its
CC complement or fragment) with a complementary nucleic acid molecule
CC obtained from the bovine cell or tissue, where hybridisation between the
CC marker nucleic acid and the complementary nucleic acid permits the
CC detection of the molecule; and (b) detecting the level or pattern of the
CC complementary nucleic acid, where the detection of the complementary
CC nucleic acid is predictive of the level or pattern of the molecule.
CC The LMFD nucleic acid is used for determining a level or pattern
CC of a molecule in a bovine cell or tissue. It is useful for genome
CC mapping, gene identification and analysis, cattle breeding, preparation
CC of constructs for use in cattle gene expression, or for genetically
CC improving cattle. The present sequence is one of the 15112 bovine
CC LMFD EST (expressed sequence tag) nucleic acids.
CC Note: The present sequence was not shown in the specification but
CC was obtained in electronic format from the USPTO web site:
CC seqdata.uspto.gov/sequence.html?docID=20020137139.
XX
SQ Sequence 201 BP; 67 A; 34 C; 23 G; 77 T; 0 other;

Query Match 27.4%; Score 95.4; DB 25; Length 201;
Best Local Similarity 76.5%; Pred. No. 1.9e-14;
Matches 117; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 174 TTGGGCTTTTCGACCACTAGACAAAATTTTTCATTTGACAGTTGAGAAATTTGTGGAGT 233
Dd |||||
193 TTGAACCTTTTCAAAAATTAGAACATAATTTTAAATTTGACAGTTGAGAAATTTGTGAGT 134
QY 234 GTTTTACATGTCTTCTTAATGAGTAGCATGTATCTTTTGCATGTATGACTTAAT 293
Dd |||||
133 GATTTTAGATGTATTTTTCATACCGTATTAAGATATCTTTTGAAGGTATCCCTTAAT 74
QY 294 AAATCCTTGAATCATAAAAAATAAAAAAT 326
Dd |||||
73 AAATCTTGACCCCTCAGTATAAAGATAAAAT 41

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 23, 2003, 09:18:57 : Search time 1304.05 Seconds
(without alignments)
10917.162 Million cell updates/sec

Title: US-09-717-321A-17

Perfect score: 348

Sequence: 1 tgaacatcactgctgcc.....cttgggaacttgaaaaaaa 348

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.om.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*

30: em.htg.hum.*

31: em.htg.inv.*

32: em.htg.other.*

33: em.htg.mus.*

34: em.htg.pla.*

35: em.htg.rod.*

36: em.htg.mam.*

37: em.htg.vrt.*

38: em.sy.*

39: em.htgo.hum.*

40: em.htgo.mus.*

41: em.htgo.other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	348	100.0	348	6	AX163753	AX163753 Sequence
c	348	100.0	1266	6	AX163738	AX163738 Sequence
3	320	92.0	217700	2	AC106124	AC106124 Rattus no
4	230.2	66.1	2281	10	BC051053	BC051053 Mus muscu
5	228.2	65.6	2319	10	BC003828	BC003828 Mus muscu
6	221	63.5	269081	2	AC068493	AC068493 Mus muscu
c	216.8	62.3	192498	2	AC105979	AC105979 Mus muscu
c	216.8	62.3	230015	2	AC132602	AC132602 Mus muscu
9	207.4	59.6	2302	9	BC050687	BC050687 Homo sapi
10	203.6	58.5	28567	9	HSAL132695	HSAL132695 Homo sapi
11	203.6	58.5	212827	9	AC009412	AC009412 Homo sapi
12	192.4	55.3	2315	9	AK054993	AK054993 Homo sapi
13	187.8	54.0	352	6	AX163754	AX163754 Sequence
14	187.4	53.9	1232	6	AX408985	AX408985 Sequence
15	187.4	53.9	1232	9	HUMPO2ST9	D25274 Homo sapien
c	176.4	50.7	87615	2	AC139405	AC139405 Homo sapi
17	176.4	50.7	137625	9	AC104663	AC104663 Homo sapi
c	175.6	50.5	240973	2	AC123247	AC123247 Rattus no
19	175.6	50.5	241048	2	AC129824	AC129824 Rattus no
c	174.8	50.2	5544	9	AF542527	AF542527 Homo sapi
c	170.6	49.0	262	11	G31709	G31709 SWSS2233 Er
c	154	44.3	455	11	G26995	G26995 human STS S
c	123	35.3	123	6	AX163737	AX163737 Sequence
24	121.4	34.9	228121	2	AC133022	AC133022 Rattus no
c	121.4	34.9	239768	2	AC112582	AC112582 Rattus no
c	121.4	34.9	245468	2	AC130746	AC130746 Rattus no
27	116.4	33.4	110816	9	AC002404	AC002404 Human Chr
28	113	32.5	1017	6	AX163751	AX163751 Sequence
c	110.6	31.8	101584	9	CNS01DS5	AL121655 BAC sequ
c	110.6	31.8	155943	9	AC012364	AC012364 Homo sapi
31	84.6	24.3	174316	2	AC022647	AC022647 Homo sapi
32	84.6	24.3	217249	9	AC009902	AC009902 Homo sapi
33	81.2	23.3	118648	9	AL138742	AL138742 Human DNA
34	81.2	23.3	131095	2	AL590071	AL590071 Homo sapi
c	79.2	22.8	170839	2	AC133467	AC133467 Mus muscu
c	75.2	21.6	237985	2	AC125754	AC125754 Rattus no
37	75.2	21.6	247478	2	AC097964	AC097964 Rattus no
38	72	20.7	156879	10	AL626786	AL626786 Mouse DNA
c	67	19.3	340701	2	AC120633	AC120633 Rattus no
40	66.8	19.2	62656	2	AC100110	AC100110 Mus muscu
c	66.8	19.2	178482	2	AC101810	AC101810 Mus muscu
42	60.8	17.5	64	6	AX522492	AX522492 Sequence
43	53.4	15.3	100697	5	AL672072	AL672072 Zebrafish
44	52.4	15.1	162	6	AX397334	AX397334 Sequence
45	52.4	15.1	1022	6	AX163752	AX163752 Sequence

ALIGNMENTS

RESULT 1
AX163753
LOCUS AX163753 348 bp DNA linear PAT 22-JUN-2001
DEFINITION Sequence 17 from Patent WO0138579.
ACCESSION AX163753
VERSION AX163753.1 GI:14544859
KEYWORDS
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Muridae; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1
AUTHORS Gould-Rothberg, B.E., Dipippo, V.A., Ramse, T.M. and Gerwein, R.W.
TITLE Method of identifying toxic agents using nsaid-induced differential

Pred. No. is the number of results predicted by chance to have a

Qy	61	AGACCTCACCACGGGACACATGTTCCGGTACCTTTGGGTCTGTGAGGTTCGTCAAG	120
Dd	297	AGACTTCACCACGGGACACATGCTTCGGTACCTTTGGGTCTGTGAGGTTCGTCAAG	238
Qy	121	CGCTAGTGCTAAACGCCGTTCTGTACAACCTAACCTCACTGCACGAACACAGTGTGGGCC	150
Dd	237	CGCTAGTGCTAAACGCCGTTCTGTACAACCTAACCTCACTGCACGAACACAGTGTGGGCC	178
Qy	181	TTTCGACCACTAGAACAACCTTTTTCAATTGCAGTTGCAGAACTGTGAGAGTGTTTTA	240
Dd	177	TTTCGACCACTAGAACAACCTTTTTCAATTGCAGTTGCAGAACTGTGAGAGTGTTTTA	118
Qy	241	CATTGATCTTTTGCATAATGCAGTAGCAGTATGTTTGCATGTATGACTTAATAAATCCT	300
Dd	117	CATTGATCTTTTGCATAATGCAGTAGCAGTATGTTTGCATGTATGACTTAATAAATCCT	58
Qy	301	TGAATCATNAAAAAAAATGCTTTTGAACCTTGAATAAAAAA 348	
Dd	57	TGAATCATNAAAAAAAATGCTTTTGAACCTTGAATAAAAAA 10	
RESULT 3		217700 bp DNA linear HTG 10-MAY-2003	
AC106124		Rattus norvegicus clone CH230-119E11, *** SEQUENCING IN PROGRESS	
LOCUS		***, 11 unordered pieces.	
DEFINITION		AC106124	
ACCESSION		AC106124_6 GI:30521557	
VERSION		HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.	
KEYWORDS		Rattus norvegicus (Norway rat)	
SOURCE		Rattus norvegicus	
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
REFERENCE		1 (bases 1 to 217700) Muzny,D.,Marie,, Metckr,M.Lee,, Abramzon,S., Adams,C., Alder,J.,, Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebech,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blythe,P., Brown,M., Bryant,N., Bulay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleeland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,K., Garcia,A., Garner,T., Garza,M., Georgievski,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulvik,S., Hume,J., Idlebird,D., Jackson,A., Jackson,J., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovac,C., Kowis,C., Kraft,C.I., Lebow,H., Levay,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseghe,H., Lozada,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartine,M., Mahmood,M., Mallory,K., Mangum,A., Manungu,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,B., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Mundasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwokeleneh,O., Okwuonu,G., Olarnpunagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfankuch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-I., Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Rem,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodney,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,,	

Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartbeyn, A., Sisson, I., Sitter, C.D., Smaj's, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, A., Trejos, Z., Umani, K., Valas, R., Vera, V., Villaseña, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, K., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GHL
Center clone name: CH230-119E11

```

----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 185725 bases at least Q40
Consensus quality: 190140 bases at least Q30
Consensus quality: 192842 bases at least Q20
Estimated insert size: 194758; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

```

```

*-----*
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of 'N', but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

1	109797: contig of 109797 bp in length
*	109897: gap of unknown length
*	109798
*	109898
*	166934: contig of 57037 bp in length
*	166935
*	167034: gap of unknown length
*	167035
*	176384: contig of 9350 bp in length
*	176385
*	176484: gap of unknown length
*	206327: contig of 29843 bp in length
*	176485
*	206427: gap of unknown length
*	206328

FEATURES	source	Location/Qualifiers	1. 217700	2. 217700	3. 217700	4. 217700	5. 217700	6. 217700	7. 217700	8. 217700	9. 217700	10. 217700	11. 217700	12. 217700	13. 217700	14. 217700	15. 217700	16. 217700	17. 217700	18. 217700	19. 217700	20. 217700	21. 217700	22. 217700	23. 217700	24. 217700	25. 217700	26. 217700	27. 217700	28. 217700	29. 217700	30. 217700	31. 217700	32. 217700	33. 217700	34. 217700	35. 217700	36. 217700	37. 217700	38. 217700	39. 217700	40. 217700	41. 217700	42. 217700	43. 217700	44. 217700	45. 217700	46. 217700	47. 217700	48. 217700	49. 217700	50. 217700	51. 217700	52. 217700	53. 217700	54. 217700	55. 217700	56. 217700	57. 217700	58. 217700	59. 217700	60. 217700	61. 217700	62. 217700	63. 217700	64. 217700	65. 217700	66. 217700	67. 217700	68. 217700	69. 217700	70. 217700	71. 217700	72. 217700	73. 217700	74. 217700	75. 217700	76. 217700	77. 217700	78. 217700	79. 217700	80. 217700	81. 217700	82. 217700	83. 217700	84. 217700	85. 217700	86. 217700	87. 217700	88. 217700	89. 217700	90. 217700	91. 217700	92. 217700	93. 217700	94. 217700	95. 217700	96. 217700	97. 217700	98. 217700	99. 217700	100. 217700	101. 217700	102. 217700	103. 217700	104. 217700	105. 217700	106. 217700	107. 217700	108. 217700	109. 217700	110. 217700	111. 217700	112. 217700	113. 217700	114. 217700	115. 217700	116. 217700	117. 217700	118. 217700	119. 217700	120. 217700	121. 217700	122. 217700	123. 217700	124. 217700	125. 217700	126. 217700	127. 217700	128. 217700	129. 217700	130. 217700	131. 217700	132. 217700	133. 217700	134. 217700	135. 217700	136. 217700	137. 217700	138. 217700	139. 217700	140. 217700	141. 217700	142. 217700	143. 217700	144. 217700	145. 217700	146. 217700	147. 217700	148. 217700	149. 217700	150. 217700	151. 217700	152. 217700	153. 217700	154. 217700	155. 217700	156. 217700	157. 217700	158. 217700	159. 217700	160. 217700	161. 217700	162. 217700	163. 217700	164. 217700	165. 217700	166. 217700	167. 217700	168. 217700	169. 217700	170. 217700	171. 217700	172. 217700	173. 217700	174. 217700	175. 217700	176. 217700	177. 217700	178. 217700	179. 217700	180. 217700	181. 217700	182. 217700	183. 217700	184. 217700	185. 217700	186. 217700	187. 217700	188. 217700	189. 217700	190. 217700	191. 217700	192. 217700	193. 217700	194. 217700	195. 217700	196. 217700	197. 217700	198. 217700	199. 217700	200. 217700	201. 217700	202. 217700	203. 217700	204. 217700	205. 217700	206. 217700	207. 217700	208. 217700	209. 217700	210. 217700	211. 217700	212. 217700	213. 217700	214. 217700	215. 217700	216. 217700	217. 217700	218. 217700	219. 217700	220. 217700	221. 217700	222. 217700	223. 217700	224. 217700	225. 217700	226. 217700	227. 217700	228. 217700	229. 217700	230. 217700	231. 217700	232. 217700	233. 217700	234. 217700	235. 217700	236. 217700	237. 217700	238. 217700	239. 217700	240. 217700	241. 217700	242. 217700	243. 217700	244. 217700	245. 217700	246. 217700	247. 217700	248. 217700	249. 217700	250. 217700	251. 217700	252. 217700	253. 217700	254. 217700	255. 217700	256. 217700	257. 217700	258. 217700	259. 217700	260. 217700	261. 217
----------	--------	---------------------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	----------

RESULT 4

BC051053

LOCUS

DEFINITION

ACCESSION
VERSTON

KEYWORDS

SOURCE

ORGANISMS

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

COMMENT


```
/lab host="DH10B"
/Note="Vector: pCMV-SPORT6"
1. 2319
/gene="Rac1"
/db_xref="synonym: D5Erd559e"
/db_xref="LocusID:19353"
/db_xref="MGI:97845"
198..776
/codon_start=1
/product="Rac1 protein"
/protein_id="AAH03828.1"
/db_xref="GI:13277918"
/db_xref="LocusID:19353"
MVDGKPNLGLWDYTAGQEDYDRPLSPQIDVFLICSLVSPASPFENVRAKWPYEV
HHCNTPILVGTGLDLDKDDKOTIEKKRKLTPITYPQGLAMAKEIGAVKYLRCAL
TORGLKTVDFDAIRVLGCPPEVKKRKKLL"
BASE COUNT 603 a 573 c 503 g 640 t
ORIGIN
Query Match 65.6%; Score 228.2; DB 10; Length 2319;
Best Local Similarity 83.7%; Pred. No. 8.5e-48;
Matches 292; Conservative 0; Mismatches 33; Indels 24; Gaps 2;
QY 21 AGAGTACCAACACTGTGATGATGATGCGCCCTCTAGACTCTACCCACGCGGACA 80
Db 1964 ACAGTACCAACACTGTGATGATGATGCGCCCTCTAGACTCTACCCACGCGGACA 2020
QY 81 CATGCTTCCGGTA-----CCTTGGGTCTGTGAGGTCTGTGCAA 119
Db 2021 GACGCTTCTGTGATGATGATGCGCCCTCTAGACTCTCTTGGGTCTGTGCAA 2080
QY 120 GCGTGTAGTGTAAAGCGTCTGTGACACCTAAGTCTGACAGTGTGAGTGTGGC 179
Db 2081 CTGCTAGTGTAAAGCGTCTGTGACACCTAAGTCTGACAGTGTGAGTGTGGC 2140
QY 180 CTTTCGACCACTAGAACAACTTTTCAATGACAGTGTGAGTGTGAGTGTGGT 239
Db 2141 CTTTCGACCACTAGAACAACTTTTCAATGACAGTGTGAGTGTGAGTGTGGT 2200
QY 240 ACATTGATCTTTGCTAATGACAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 299
Db 2201 ACATTGATCTTTGCTAATGACAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2260
QY 300 TTGAATCATATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 348
Db 2261 TTGAATCATATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 2309
RESULT 6
AC068493
LOCUS
DEFINITION Mus musculus clone RP23-76K1 strain C57BL6/J, WORKING DRAFT
SEQUENCE, 26 unordered pieces.
AC068493
VERSION AC068493.10 GI:15148081
KEYWORDS HTG: HTGS PHASE1; HTGS DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 269081)
Montgomery K.T., Grills G., Han J., Lee E., Long J., Pomerantz R.,
Toshikhes I.P., Shim C., Decker J., Thomas E., Perera A.,
Gordon M., Goltz J.S. and Kucherlapati R.
High Throughput Mouse Sequencing
Unpublished
2 (bases 1 to 269081)
Montgomery K.T., Grills G., Han J., Lee E., Long J., Pomerantz R.,
Toshikhes I.P., Shim C., Decker J., Thomas E., Perera A.,
Gordon M., Goltz J.S. and Kucherlapati R.
Direct Submission
Submitted (03-MAY-2000) Department of Molecular Genetics, Albert
```

COMMENT

Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
Bronx, NY 10461, USA
On Aug 11, 2001 this sequence version replaced gi:14993654.
-----Genome Center
Center: Harvard Partners Genome Center
Center Code: HPGC
Web site: <http://www.hpcgg.org/Sequence/mouse.html>
Contact: hpgc@mendel.mgh.harvard.edu
-----Summary Statistics
Center project name: ABN
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 100%
*Consensus quality: 256302 at least Q20
*Consensus quality: 254124 at least Q30
*Consensus quality: 250276 at least Q40
Estimated insert size: agarose-FP - N/A
**Estimated insert size: 268581 - sum-of-contigs
Quality coverage: agarose-FP - N/A
Quality coverage: 6.2 x in Q20 bases; sum-of-contigs estimation

* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 51730: contig of 51730 bp in length
* 51731 51750: gap of unknown length
* 51751 82423: contig of 30673 bp in length
* 82424 82443: gap of unknown length
* 82444 118158: contig of 35715 bp in length
* 118159 118178: gap of unknown length
* 118179 139441: contig of 21263 bp in length
* 139442 139461: gap of unknown length
* 139462 160638: contig of 21177 bp in length
* 160639 160658: gap of unknown length
* 160659 177611: contig of 16953 bp in length
* 177612 177631: gap of unknown length
* 177632 190991: contig of 13360 bp in length
* 190992 191011: gap of unknown length
* 191012 200128: contig of 9117 bp in length
* 200129 200148: gap of unknown length
* 200149 211882: contig of 11734 bp in length
* 211883 211902: gap of unknown length
* 211903 220773: contig of 8871 bp in length
* 220774 220793: gap of unknown length
* 220794 229643: contig of 8850 bp in length
* 229644 229663: gap of unknown length
* 229664 238339: contig of 8676 bp in length
* 238340 238359: gap of unknown length
* 238360 244848: contig of 6489 bp in length
* 244849 244868: gap of unknown length
* 244869 249725: contig of 4857 bp in length
* 249726 252523: contig of 2778 bp in length
* 252524 252543: gap of unknown length
* 252544 257520: contig of 4977 bp in length
* 257521 257540: gap of unknown length
* 257541 260396: contig of 2856 bp in length
* 260397 260416: gap of unknown length
* 260417 262709: contig of 2293 bp in length
* 262710 262729: gap of unknown length
* 262730 263934: contig of 1205 bp in length
* 263935 264196: gap of unknown length
* 264197 264216: gap of unknown length
* 264217 264629: contig of 413 bp in length
* 264630 264649: gap of unknown length
* 264650 265608: contig of 959 bp in length
* 265609 265628: gap of unknown length
* 265629 266935: contig of 1307 bp in length

Query Match 63.5%; Score 221; DB 2; Length 269081;
Best Local Similarity 82.9%; Pred. No. 7.3e-46;

Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Bouhgalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,

JOURNAL

Submitted (01-SEP-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

COMMENT

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
Contact: submissions@wustl.edu
----- Project Information -----
Center project name: M_BA0477C13

----- Summary Statistics -----
Sequencing vector: M13, 0%
Sequencing vector: plasmid, 100%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 225083 bases at least Q40
Consensus quality: 225730 bases at least Q30
Consensus quality: 226267 bases at least Q20
Insert size: 202000; agarose-fp
Insert size: 228405; sum-of-contigs
Quality coverage: 13.11 in Q20 bases; agarose-fp
Quality coverage: 9.64 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 10 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 710: contig of 710 bp in length
711 810: gap of unknown length
811 1816: contig of 17506 bp in length
1817 18416: gap of unknown length
18417 57380: contig of 38964 bp in length
57381 57480: gap of unknown length
57481 118173: contig of 60693 bp in length
118174 118273: gap of unknown length
118274 201518: contig of 83245 bp in length
201519 201618: gap of unknown length
201619 202801: contig of 1183 bp in length
202802 202901: gap of unknown length
202902 204714: contig of 1813 bp in length
204715 204814: gap of unknown length
204815 208091: contig of 3277 bp in length
208092 208191: gap of unknown length
208192 213082: contig of 4891 bp in length
213083 213183: gap of unknown length
213183 230015: contig of 16833 bp in length.

FEATURES

source

1. .230015
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP23-477C13"

1. .710
/note="assembly_name:Contig1"
811. .1816
/note="assembly_name:Contig10"
18417. .57380
/note="assembly_name:Contig11"
57481. .118173
/note="assembly_name:Contig12"
118274. .201518
/note="assembly_name:Contig13"
201619. .202801
/note="assembly_name:Contig3"
202902. .204714
/note="assembly_name:Contig6"

misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature

misc_feature 204815. .208091
/note="assembly_name:Contig7"
misc_feature 208192. .213082
/note="assembly_name:Contig8"
misc_feature 213183. .230015
/note="assembly_name:Contig9"

BASE COUNT 63025 a 50551 c 50278 g 65213 t 948 others
ORIGIN

Query Match 62.3%; Score 216.8; DB 2; Length 230015;
Best Local Similarity 86.1%; Pred. No. 8.6e-45;
Matches 272; Conservative 0; Mismatches 22; Indels 22; Gaps 2;

QY 29 AACACTGTCATGTGATTCGCGCCCTCTAGACCTCACCCACGGGACACATGCTTC 88
DB 109281 AACACTGTCATGTGACTAATGCTG-CCCTCTAGACCTCACTCAGGACAGCGCTTC 109223
QY 89 CGGTA-----CCTTGGGCTCTGAGGTTCTGCAAGCGCTAGT 127
DB 109222 CTGCTGGCTCTGCTAGAGATGTTCTCTGGGGTCTGTGAGGTTCTGCAACTGCTAGT 109163
QY 128 GCTAACGGCGTCTGTGACAACTTAACCTCACTGCGAAGAACACAGTCTTGGGCTTTCGAC 187
DB 109162 GCTAATGCTGCTGTGACAACTTAACCTCACTGCGAAGAACACAGTCTTGGGCTTTCGAC 109103
QY 188 CACTAGACAAACTTTTTTCAATTGACAGTTGCGAATTTGTGGAGTGTTTTACATTGAT 247
DB 109102 CACTAGACAAACTTTTTTCAATTGACAGTTGCGAATTTGTGGAGTGTTTTACATTGAT 109043
QY 248 CTTTGTCTAATGACAGTTGACAGTATGTTTGTGATGATGACATTAATAATCCTTGAATCA 307
DB 109042 CTTTGTCTAATGACAGTTGACAGTATGTTTGTGATGATGACATTAATAATCCTTGAATCA 108983

QY 308 TAAAAAATAAAAAA 323
DB 108982 TAAAAAATAAAAAA 108967

RESULT 9

BC050687

LOCUS

DEFINITION

BC050687

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

COMMENT

BC050687 2302 bp mRNA linear PRI 11-APR-2003
Homo sapiens, ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1), clone MGC:60264
IMAGE:6149377, mRNA, complete cds.

BC050687.1 GI:29792301

MGC.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 2302)

Strausberg, R.

Direct Submission

Submitted (08-APR-2003) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgapbs@mail.nih.gov

Tissue Procurement: ATCC/DCTD/PTP

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Sequencing Group at the Stanford Human Genome

Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: <http://www.shgc.stanford.edu>

Contact: (Dickson, Mark) mcdpaxil@stanford.edu

Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,

R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAK Plate: 110 Row: k Column: 14
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 9845510.

FEATURES
source
 1..2302
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:60264 IMAGE:6149377"
 /tissue_type="Skin, melanotic melanoma."
 /clone_lib="NIH_MGC_72"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 189..767
 /codon_start=1
 /product="ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1)"
 /protein_id="AAH50687.1"
 /db_xref="GI:29792302"
 /db_xref="LocusID:5879"
 /translation="MQAIKCVVGDGAVGKTCLLISYTNAPFGEYIPTVFDNYSANV
 MVDGKPNVLGLWDTAGQEDYDLRLPLSYPTQDVELICFSLVSPASFENVRKAWPEVR
 HHCENPILIVGKTLDRDDKDTIEKLEKRLPTITYPQGLAKAKEIGAVKYLECSAL
 TQGLKTVFDEAIRAVLCPPPVKKRKKCLLL"
 617 a 524 c 498 g 663 t

BASE COUNT
 617 a 524 c 498 g 663 t

ORIGIN
 Query Match 59.6%; Score 207.4; DB 9; Length 2302;
 Best Local Similarity 82.5%; Pred. No. 1.8e-42;
 Matches 283; Conservative 0; Mismatches 36; Indels 24; Gaps 3;

QY 4 AACATCACTCTCTTGCAGAGTACCAACACTGTCATGTGATTGATCGCCCTCTAGA 63
Db 1961 AACATCACTCTCTTGCAGATTACCGACACTGTCATGACCAACTGA-CCCTCTTTA 2019

QY 64 CCTCACCCAGCGGACACATGTTCCGGTA-----CCTTTGGGT 102
Db 2020 CCTCGCCCGCGGACACAGCGCTCTGTAGTCGTTTGCTATTGATGTTCTTTGGGT 2079

QY 103 CTGTGAGGTTCTGTCAA--GCGTAGTGTCTAAGCGCTTCTGTACAACTCACTCTGG 160
Db 2080 CTGTGAGGTTCTGTAACTGTGTAGTGTGAGGATGTTCTGTACAACTTAACCTCTGG 2139

QY 161 CAAGAACACAGTGTGGGCTTTTCGACCACACTGACAACTTTTTCATTTGACAGTTGC 220
Db 2140 CGAAGATACAGGTGGAGCCCTCAGCCACTACACAGAAATTTTAAATTGACAGTTGC 2199

QY 221 AGAATTGTGGAGTGTTTTACATTGATCTTTTCTAATGAGTTAGCAGTATGTTTTCGA 280
Db 2200 AGAATTGTGGAGTGTTTTACATTGATCTTTTCTAATGAGTTAGCAGTATGTTTTCGA 2259

QY 281 TGTATGACTTAATAATCTCTGAATCATATAAAAAAAAAAAAA 323
Db 2260 TGTATGACTTAATAATCTTGAATCATATAAAAAAAAAAAAA 2302

RESULT 10
HSAL32695
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
 HSAL32695
 Homo sapiens rac1 gene.
 AJ132695
 AJ132695.5 GI:8574037
 alternative splicing; Alu; AluJ; AluSg1; AluSp; AluX; AT-rich; CT-rich; MIR; rac1 gene; rac1 protein; Rac1b protein; repetitive sequence;
 Homo sapiens (human)

SOURCE
ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1
 Matos, P., Skaug, J., Marques, B., Beck, S., Verissimo, F., Gespach, C., Boavida, M.G., Scherer, S.W. and Jordan, P.
 Small GTPase Rac1: structure, localization, and expression of the

human gene
 Biochem. Biophys. Res. Commun. 277 (3), 741-751 (2000)
 MEDLINE 20517245
 PUBMED 11062023
 2 (bases 1 to 28567)
 Jordan, P.
 Direct Submission
 Submitted (03-FEB-1999) Jordan P., Centro de Genetica Humana, Laboratorio de Oncobiologia, Instituto Nacional de Saude 'Dr. Ricardo Jorge', Avenida Padre Cruz, 1649-016 Lisboa, PORTUGAL
 On Jun 20, 2000 this sequence version replaced gi:7248282.
 Related sequence AL638561.
 Location/Qualifiers
 1..28567
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="7"
 /map="7p22"
 /clone="H..NH0425A05, RPCI 11"
 1..109
 /rpt_family="AluSg"
 125..422
 /rpt_family="AluSc"
 441..742
 /rpt_family="AluSg"
 743..28117
 /gene="rac1"
 743..1226
 /gene="rac1"
 911..983
 /rpt_family="GC-rich"
 936
 /gene="rac1"
 /note="alternative transcription start site"
 1026
 /genes="rac1"
 /note="alternative transcription start site"
 1037
 /genes="rac1"
 /note="alternative transcription start site"
 1053
 /gene="rac1"
 /note="alternative transcription start site"
 1106..1175
 /rpt_family="GC-rich"
 join(1227..1261,13701..13772,18413..18530,22804..22860,24268..24330,26010..26169,26458..26588)
 /gene="rac1"
 /note="alternative"
 /codon_start=1
 /product="Rac1b protein"
 /protein_id="CAA10733.6"
 /db_xref="GI:8574039"
 /db_xref="SPTREMBL:O95501"
 /translation="MQAIKCVVGDGAVGKTCLLISYTNAPFGEYIPTVFDNYSANV
 MVDGKPNVLGLWDTAGQEDYDLRLPLSYPTQDVELICFSLVSPASFENVRKAWPEVR
 GLVSPASFENVRKAWPEVRHHCENPILIVGKTLDRDDKDTIEKLEKRLPTITYPQGLAKAKEIGAVKYLECSAL
 TQGLKTVFDEAIRAVLCPPPVKKRKKCLLL"
 join(1227..1261,13701..13772,18413..18530,24268..24330,26010..26169,26458..26588)
 /gene="rac1"
 /codon_start=1
 /product="Rac1 protein"
 /protein_id="CAB53579.5"
 /db_xref="GI:8574038"
 /translation="MQAIKCVVGDGAVGKTCLLISYTNAPFGEYIPTVFDNYSANV
 HVDGKPNVLGLWDTAGQEDYDLRLPLSYPTQDVELICFSLVSPASFENVRKAWPEVR
 HHCENPILIVGKTLDRDDKDTIEKLEKRLPTITYPQGLAKAKEIGAVKYLECSAL
 TQGLKTVFDEAIRAVLCPPPVKKRKKCLLL"
 <1227..1261
 /gene="rac1"
 /number=1

exon

Repeat Type	Repeat ID	Repeat Sequence	Repeat Family	Repeat Length	Repeat Count	Repeat Position	Repeat Orientation	Repeat Annotation
intron	1262..13700	/gene="rac1"	/number=1	1390..1419	1752..1793	1794..2005	2140..2171	2666..2972
repeat_region	1390..1419	/rpt family="GC_rich"	1752..1793	1794..2005	2140..2171	2666..2972	1390..1419	1752..1793
repeat_region	1752..1793	/rpt family="FLAM"	1794..2005	2140..2171	2666..2972	1390..1419	1752..1793	1794..2005
repeat_region	1794..2005	/rpt family="AluJo"	2140..2171	2666..2972	1390..1419	1752..1793	1794..2005	2140..2171
repeat_region	2140..2171	/rpt family="Simple_repeat"	2666..2972	1390..1419	1752..1793	1794..2005	2140..2171	2666..2972
repeat_region	2666..2972	/rpt family="AluSx"	1390..1419	1752..1793	1794..2005	2140..2171	2666..2972	1390..1419
repeat_region	1390..1419	/rpt family="AluSx"	1752..1793	1794..2005	2140..2171	2666..2972	1390..1419	1752..1793
repeat_region	1752..1793	/rpt family="Tn"	1794..2005	2140..2171	2666..2972	1390..1419	1752..1793	1794..2005
repeat_region	1794..2005	/rpt family="AluJb"	2140..2171	2666..2972	1390..1419	1752..1793	1794..2005	2140..2171
repeat_region	2140..2171	/rpt family="AluSx"	2666..2972	1390..1419	1752..1793	1794..2005	2140..2171	2666..2972
repeat_region	2666..2972	/rpt family="AluSx"	1390..1419	1752..1793	1794..2005	2140..2171	2666..2972	1390..1419
repeat_region	1390..1419	/rpt family="LTR33A"	1752..1793	1794..2005	2140..2171	2666..2972	1390..1419	1752..1793
repeat_region	1752..1793	/rpt family="AluSx"	1794..2005	2140..2171	2666..2972	1390..1419	1752..1793	1794..2005
repeat_region	1794..2005	/rpt family="AluSx"	2140..2171	2666..2972	1390..1419	1752..1793	1794..2005	2140..2171
repeat_region	2140..2171	/rpt family="AluSx"	2666..2972	1390..1419	1752..1793	1794..2005	2140..2171	2666..2972
repeat_region	2666..2972	/rpt family="AluSx"	1390..1419	1752..1793	1794..2005	2140..2171	2666..2972	1390..1419
repeat_region	1390..1419	/rpt family="AluSx"	1752..1793	1794..2005	2140..2171	2666..2972	1390..1419	1752..1793
repeat_region	1752..1793	/rpt family="AluSx"	1794..2005	2140..2171	2666..2972	1390..1419	1752..1793	1794..2005
repeat_region	1794..2005	/rpt family="AluSx"	2140..2171	2666..2972	1390..1419	1752..1793	1794..2005	2140..2171
repeat_region	2140..2171	/rpt family="AluSx"	2666..2972	1390..1419	1752..1793	1794..2005	2140..2171	2666..2972
repeat_region	2666..2972	/rpt family="AluSx"	1390..1419	1752..1793	1794..2005	2140..2171	2666..2972	1390..1419
repeat_region	1390..1419	/rpt family="AluSx"	1752..1793	1794..2005	2140..2171	2666..2972	1390..1419	1752..1793
repeat_region	1752..1793	/rpt family="AluSx"	1794..2005	2140..2171	2666..2972	1390..1419	1752..1793	1794..2005
repeat_region	1794..2005	/rpt family="AluSx"	2140..2171	2666..2972	1390..1419	1752..1793	1794..2005	2140..2171
repeat_region	2140..2171	/rpt family="AluSx"	2666..2972	1390..1419	1752..1793	1794..2005	2140..2171	2666..2972
repeat_region	2666..2972	/rpt family="AluSx"	1390..1419	1752..1793	1794..2005	2140..2171	2666..2972	1390..1419
repeat_region	1390..1419	/rpt family="AluSx"	1752..1793	1794..2005	2140..2171	2666..2972	1390..1419	1752..1793
repeat_region	1752..1793	/rpt family="AluSx"	1794..2005	2140..2171	2666..2972	1390..1419	1752..1793	1794..2005
repeat_region	1794..2005	/rpt family="AluSx"	2140..2171	2666..2972	1390..1419	1752..1793	1794..2005	2140..2171
repeat_region	2140..2171	/rpt family="AluSx"	2666..2972	1390..1419	1752..1793	1794..2005	2140..2171	2666..2972

RESULT 11
AC009412 212827 bp DNA linear PRI 07-NOV-2001
LOCUS Homo sapiens BAC clone RP11-425P5 from 7, complete sequence.
DEFINITION AC009412
ACCESSION AC009412
VERSION AC009412.6 GI:14190769
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 212827)
AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 212827)
AUTHORS Hou, S., Maupin, R., Haakenson, W., Gregory, S. and Belter, R.
TITLE The sequence of Homo sapiens BAC clone RP11-425P5
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 212827)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (21-AUG-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 212827)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (23-MAY-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 212827)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (07-NOV-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On May 23, 2001 this sequence version replaced gi:13431187.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics

Center project name: H_RH0425P05

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see
http://www.nhgri.nih.gov/DIR/GRB/CHR7, send
mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male

donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)
VECTOR: pBACE3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is CTD-2195P2, 200 bp overlapi. Actual start of this clone is at base position 1 of RP11-425P5; actual end is at base position 212633 of RP11-425P5.
Location/Qualifiers
1. 212827
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="7"
/map="7"
/clone="RP11-425P5"
/clone_lib="RPCI-11"
1623..1696
/rpt_family="MIR"
1698..2006
/rpt_family="Alu"
2024..2337
/rpt_family="Alu"
3006..3093
/rpt_family="L1"
3453..3552
/rpt_family="L2"
4924..5046
/rpt_family="ERV1"
5318..5693
/note="similar to EST BF996294 (NID:gil2402607)"
5580..5845
/rpt_family="Alu"
5925..6277
/rpt_family="MALR"
6842..7041
/note="similar to EST BF369507 (NID:gil1331532)"
7364..7661
/rpt_family="Alu"
8029..8048
/rpt_family="(CAAAA)n"
8379..8678
/rpt_family="Alu"
8560..8762
/note="similar to EST AA699398 (NID:g2702592) z140a03.s1"
8720..9031
/rpt_family="Alu"
9083..9560
/note="similar to EST AA633603 (NID:g25556817) ae66b02.s1"
9598..9899
/rpt_family="Alu"
9871..10098
/note="similar to EST AA699398 (NID:g2702592) z140a03.s1"
10746..11044
/rpt_family="L1"
11075..11101
/rpt_family="AT-rich"
11102..11372
/rpt_family="Alu"
11411..11465
/rpt_family="GA-rich"
11552..11644
/rpt_family="L2"
11685..12193
/rpt_family="L1"
12194..12486
/rpt_family="Alu"
12487..13003
/rpt_family="L1"

	repeat_region	13004. .13305 /rpt_family="Alu"	
	repeat_region	13306. .13501 /rpt_family="L1"	
	repeat_region	13503. .13892 /rpt_family="L1"	
	misc_feature	13893. .14224 /note="similar to EST BE061102 (NID:g8405752)"	
	repeat_region	14244. .14544 /rpt_family="Alu"	
	repeat_region	14546. .14828 /rpt_family="Alu"	
	repeat_region	15091. .15189 /rpt_family="MIR"	
	repeat_region	15672. .16449 /rpt_family="Achobo"	
	repeat_region	16640. .16844 /rpt_family="Alu"	
	repeat_region	16917. .16959 /rpt_family="Achobo"	
	repeat_region	16960. .17287 /rpt_family="Alu"	
	repeat_region	17288. .17572 /rpt_family="Achobo"	
	misc_feature	17610. .17938 /note="similar to EST BF996427 (NID:g12402750)"	
	repeat_region	17793. .18068 /rpt_family="Alu"	
	repeat_region	18251. .18537 /rpt_family="Alu"	
	misc_feature	18485. .18884 /note="similar to EST BE062347 (NID:g8406997)"	
	misc_feature	18661. .18809 /note="similar to EST BG186544 (NID:g13708231)"	
	repeat_region	18957. .19160 /rpt_family="L1"	
	repeat_region	19314. .19528 /rpt_family="L1"	
	repeat_region	19819. .20135 /rpt_family="Alu"	
	repeat_region	21101. .21269 /rpt_family="L1"	
	repeat_region	21882. .22203 /rpt_family="L1"	
	misc_feature	22429. .22797 /note="similar to EST BF993478 (NID:g12399801)"	
	repeat_region	22717. .23007 /rpt_family="L1"	
	misc_feature	22993. .23191 /note="similar to EST BF767044 (NID:g12115035)"	
	repeat_region	23167. .23471 /rpt_family="Alu"	
	misc_feature	23511. .23930 /note="similar to EST AW105611 (NID:g6076346) xd49g02.xl"	
	repeat_region	25387. .25677 /rpt_family="Alu"	
	Query Match	58.5%; Score 203.6; DB 9; Length 212827;	
	Best Local Similarity	78.8%; Pred.No.2e-41; Indels 24; Gaps 3;	
	Matches 290;	Conservative 0; Mismatches 0;	
Qy	4 AACATCACTGTTGCCAGACTACAACTGTCTAATGATTGATGC GCCGCCCTCTAGA 63		
Db	175810 AACATCACTGTTGCCAGACTACCGACACTGTCACTTGACCATACTGA-CCTCTTTTA 175868		
Qy	64 CCTCACCCACGGCACAATGCTTCCGGTA-----CCTTTGGGT 102		
Db	175869 CCTCGCCACGGGACACACGCCTCCTGTAGTCGTTGCCCTATTGANGTTCCTTTGGT 175928		
Qy	103 CTGTGAGGTTCTGTCAA--GGCTAGTGCTAACGCCGTTTCTGTACAACTCACTGCG 160		
Db	175929 CTGTGAGTCTGTAAACTGTGCTAGTGTGACGATGTTCTGTACAACTTAACCTCACTGG 175988		

```

QY 64 CCTCACCACGGGACACATGCTTCGGTA-----CCTTTGGGT 102
Db 2048 CCTCGCCACGGGACACACGCTCTCTAGTCGCTTTCCTTATGATGTTCTTTGGGT 2107
QY 103 CTGTGAGTTCTGTCAA--GCCTAGTGCTAACGCCGTTCTGTACAACCTAACTCACTGG 160
Db 2108 CTGTGAGTTCTGTAACTGTCTAGTCGATGTTCTGTACAACCTAACTCACTGG 2167
QY 161 CAAGAACACAGTGTGGCCTTTCGACCACATAGAACAACTTTTCAATTGACAGTTGC 220
Db 2168 CGAGNATACGCTGGGACCTTCAGCCACTACAACAGAAATTTTAAATTCACAGTTGC 2227
QY 221 AGAATTGTGAGTGTTTTATCATTCATCTTTGTAAATGCAGTTAGCAGTATGTTTGA 280
Db 2228 AGAATTGTGAGTGTTTTATCATTCATCTTTGTAAATGCAGTTAGCAGTATGTTTGA 2287
QY 281 TGTATGACTTAATAATCCTTGAATCAT 308
Db 2288 TGTATGACTTAATAATCCTTGAATCAT 2315

RESULT 13
AX163754
LOCUS AX163754 352 bp DNA linear PAT 22-JUN-2001
DEFINITION Sequence 18 from Patent WO0138579.
ACCESSION AX163754
VERSION AX163754.1 GI:14544860
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Gould-Rorhberg,B.E., Dippio,V.A., Rameeh,T.M. and Gerwein,R.W.
TITLE Method of identifying toxic agents using nsaid-induced differential
gene expression in liver
JOURNAL Patent: WO 0138579-A 18 31-MAY-2001;
Curagen Corporation (US)
FEATURES
Location/Qualifiers
source 1..352
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 86 a 77 c 72 g 117 t
ORIGIN
Query Match 54.0%; Score 187.8; DB 6; Length 352;
Best Local Similarity 77.6%; Pred. No. 1.7e-37;
Matches 274; Conservative 0; Mismatches 57; Indels 22; Gaps 3;
QY 17 TGCAGAGTACCAACACTGTCATGTCATGATGATCGCGCCCTTAGACCTCACCCACGG 76
Db 1 TGCAGATTAACGACACTGTCATTGACCAATACG-ACCCTCTTTACCTCGCCACGG 59
QY 77 GACACATGCTTCG-----GTACCTTTGGGTCTGTGAGTTCTGTC 117
Db 60 GACACGCTCTGTCGTGCTGCTTGCCTATGATGTTCCCTTTGGGTCTGTGAGTTCTGTA 119
QY 118 AA--GGCTAGTCTAACCGGTTCTGTACACCTAATCACTGCGAAGAACACAGTGT 175
Db 120 AACTGTGCTAGTGTGACGATGTTCTGTACAACTTAATCACTGCGGAGAAATACAGCGT 179
QY 176 GGGCCCTTCGACCCTAGAACAACTTTTCAATTGACAGTTGCAATTTGGAGTGT 235
Db 180 GGACCTTCAGCACATACACAGAAATTTTAAATTCAGATTCAGAAATTTGGAGTGT 239
QY 236 TTTTACATGATCTTTTGTAAATGAGTATGATGTTTGTGATGATGATTAATAA 295
Db 240 TTTTACATGATCTTTTGTAAATGAGTATGATGTTTGTGATGATGATTAATAA 299
QY 296 ATCCTTGAATCATAAAAAATAAATAAATGCTTTTGAATCTGAAAAA 348

```

```

Db 300 ATCCTTGAATCATACGACTGGTAATACTAGTGTCTTTTGAGACTGTGATGAACA 352

RESULT 14
AX408985
LOCUS AX408985 1232 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 1632 from Patent WO0229103.
ACCESSION AX408985
VERSION AX408985.1 GI:21441690
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Alvares,C., Horne,D., Peres-da-Silva,S. and Vockley,J.G.
TITLE Gene expression profiles in liver cancer
JOURNAL Patent: WO 0229103-A 1632 11-APR-2002;
GENE LOGIC INC (US)
FEATURES
Location/Qualifiers
source 1..1232
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/notes="EMBL/GenBank Accession No. D25274"
BASE COUNT 329 a 240 c 251 g 411 t 1 others
ORIGIN
Query Match 53.9%; Score 187.4; DB 6; Length 1232;
Best Local Similarity 80.9%; Pred. No. 2.2e-37;
Matches 263; Conservative 1; Mismatches 37; Indels 24; Gaps 3;
QY 4 AACATCACTGCTTTCGACAGATCAACACTGTCATGATGATGATGCGCCCTCTAGA 63
Db 905 AACATCACTGCTTTCGACAGATCAACACTGTCATGATGATGATGCGCCCTCTAGA 963
QY 64 CCTCACCACGGGACACATGCTTCGGTA-----CCTTTGGGT 102
Db 964 CCTCGCCACGGGACACACGCTCTCTAGTCGCTTTCCTTATGATGTTCTTTGGGT 1023
QY 103 CTGTGAGTTCTGTCAA--CGCTAGTGTAAAGCCGCTTGTACAACTCACTCACTGG 160
Db 1024 CTGTGAGTTCTGTAAACTGTGCTAGTGTGACAGTGTCTGTACAACTCACTCACTGG 1083
QY 161 CAAGAACACAGTGTGGGCTTTTCGACCACTAGAACAACTTTTCAATTGACAGTTGC 220
Db 1084 CGAGATACAGCTGGGACCTTCAGCCACTACACAGAAATTTTAAATTCAGAGTTGC 1143
QY 221 AGAATTGTGAGTGTTTTACATGATCTTTGCTAATGAGTTAGCAGTATGTTTGA 280
Db 1144 AGAATTGTGAGTGTTTTACATGATCTTTGCTAATGCAATTAGCAATATGTTTGA 1203
QY 281 TGTATGACTTAATAATCCTTGAAT 305
Db 1204 TGTATGACTTAATAATCCTTGAAT 1228

RESULT 15
HUMPO2ST9
LOCUS HUMPO2ST9 1232 bp mRNA linear PRI 11-MAR-1998
DEFINITION Homo sapiens mRNA, clone:PO2ST9.
ACCESSION D25274
VERSION D25274.1 GI:464185
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Ikeda,J., Hadano,S., Nagayama,T., Tomiyasu,H., Wakasa,K. and
TITLE Isolation and characterization of 21 novel expressed DNA sequences
from the distal region of human chromosome 4p

```

COMMENT

Fax: 0463-91-4110.

FEATURES

```
1. .1232 "/organism="Homo sapiens"  
/mol_type="mRNA"  
/db_xref="taxon:9606"  
/chromosome="4"  
/clone="p02Str9"  
/tissue_type="brain striatum"  
1214. .1219
```

polvA signal

polyA-site

329 a	240 c	251 g	411 t	1 others
-------	-------	-------	-------	----------

ORIGIN

Query Match	53.9%;	Score 187.4;	DB 9;	Length 1232;	
Best Local Similarity	80.9%;	Pred. No. 2.2e-37;			
Matches 263;	Conservative	1;	Mismatches 37;	Indels 24;	Gaps 3;
4	AACATCACTGCTTCGCCAGGTACCAACACTCTCATGTGATTGATGCGCCGCCCTCTAGA	63			
905	AACATCACTGCTTCGCCAGATTACGACACTGTCATCTGACCAACTACTGA-CCCTCTTTA	963			
64	CCTCACCACCGCGGACACATGCTTCGGTA-----CCTTTGGGT	102			
964	CCTCGCCACCGCGGACACACGCTCCTGTAGTCGGTTTGCCTATTGATGTCCTTTGGGT	1023			
103	CTGTGAGGTTCTGTCAA--GCGCTAGTGTCTAACGCGGTTCTGTACACCTACTCACTGG	160			
1024	CTGTGAGGTTCTGTAAACTGTCTAGTGTGACGATGTTCTGTACAACTTAACCTCACTGG	1083			
161	CAAGAACACAGTGTGGGCGCTTCGACCACTAGAACAACTTTTTTCAATTGACAGTTGC	220			
1084	CGAGAAATACAGCGTGGGACCCCTTCAGCCACTACACAGAAATTTTTTAATTGACAGTTGC	1143			
221	AGAAATGTGCGAGTGTTTTTACATTGATCTTTTGCTAAATGCAGTATGCACTGTTTGC	280			
1144	AGAAATGTGCGAGTGTTTTTACATTGATCTTTTGCTAAATGCAGTATGCTGCA	1203			
281	TGATGACTTAATAAATCCTTGAAT	305			
1204	TGATGACTTAATAAATCCTTGGAA	1228			

Search completed: November 23, 2003, 14:02:04
Job time : 1307.05 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 23, 2003, 10:50:08 ; Search time 26.5359 Seconds
(without alignments)
5788.425 Million cell updates/sec

Title: US-09-717-321A-17
Perfect score: 348
Sequence: 1 tgaacatcactgtctgcc.....ctttggaacttgaaaaaaa 348

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	52.4	15.1	605	3	US-09-385-982-483
C 2	43	12.4	38844	4	US-09-734-675-3
C 3	40.8	11.7	5852	1	US-07-867-106-2
C 4	39.2	11.3	859	1	US-08-345-756-8
C 5	39.2	11.3	859	1	US-08-625-198-8
C 6	39.2	11.3	2039	1	US-08-345-756-5
C 7	39.2	11.3	2039	1	US-08-625-198-5
C 8	38.8	11.1	760	4	US-09-205-258-232
C 9	38.8	11.1	2218	4	US-09-205-258-103
C 10	36.8	10.6	1217	3	US-09-277-716-17
C 11	36.8	10.6	1217	4	US-09-609-161B-17
C 12	36.4	10.5	1454	4	US-08-745-995A-31
C 13	36.4	10.5	1454	4	US-08-745-995A-33
C 14	36.4	10.5	1454	4	US-09-005-352-31
C 15	36.4	10.5	1454	4	US-09-005-352-33
C 16	36.4	10.5	1584	4	US-08-745-995A-1
C 17	36.4	10.5	1584	4	US-08-745-995A-3
C 18	36.4	10.5	1584	4	US-09-005-352-1
C 19	36.4	10.5	1584	4	US-09-005-352-3
C 20	36.4	10.5	1838	4	US-08-745-995A-13
C 21	36.4	10.5	1838	4	US-08-745-995A-15
C 22	36.4	10.5	1838	4	US-09-005-352-13
C 23	36.4	10.5	1838	4	US-09-005-352-15
C 24	36	10.3	1882	3	US-09-370-253-1
C 25	36	10.3	3736	2	US-08-480-473B-1
C 26	36	10.3	3736	3	US-08-915-213-1
C 27	36	10.3	3736	3	US-09-148-547-1

```

28 36 10.3 3736 3 US-09-235-217-1 Sequence 1, Appli
29 36 10.3 3736 4 US-09-383-581-1 Sequence 1, Appli
30 36 10.3 3736 5 PCT-US96-10251-1 Sequence 1, Appli
31 35.8 10.3 566 4 US-09-105-542A-1 Sequence 1, Appli
32 35.6 10.2 501 4 US-09-601-198-170 Sequence 170, App
33 35.6 10.2 991 3 US-08-924-747-25 Sequence 25, Appli
34 35.6 10.2 991 3 US-09-247-373B-25 Sequence 25, Appli
35 35.6 10.2 991 3 US-09-296-715-25 Sequence 19, Appli
36 35.6 10.2 9829 4 US-09-322-478-19 Sequence 17, Appli
37 35.6 10.2 12286 4 US-09-322-478-17 Sequence 2, Appli
38 35.4 10.2 1571 5 PCT-US94-12912-2 Sequence 2, Appli
39 35.2 10.1 615 4 US-09-105-542A-2 Sequence 1, Appli
40 35 10.1 580073 4 US-08-545-528D-1 Sequence 11, Appli
41 34.8 10.0 13865 3 US-09-009-217-11 Sequence 11, Appli
42 34.8 10.0 13865 3 US-09-009-656-11 Sequence 66, Appli
43 34.6 9.9 664 4 US-09-904-615-66 Sequence 17, Appli
44 34.6 9.9 1596 3 US-09-146-950-17 Sequence 1, Appli
45 34.6 9.9 1929 3 US-09-146-950-1 Sequence 1, Appli

```

ALIGNMENTS

```

RESULT 1
US-09-385-982-483/c
; Sequence 483, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: PRODUCTS: 11
; FILE REFERENCE: CCDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1993-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 483
; LENGTH: 605
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(605)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-483

```

```

Query Match 15.1%; Score 52.4; DB 3; Length 605;
Best Local Similarity 81.1%; Pred. No. 3.7e-05;
Matches 73; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

```

```

QY 4 AACATCACTGCTTGGCAGAGTACCACTGATTCATGTCGCGCCCTCTAGA 63
|||||
Db 162 AACATCACTGCTTGGCAGATTACCGACACTGTCCTTGACCAATCTGA 104
|||||
QY 64 CCTCACCGCGGACACATGCTTCCGGTA 93
|||||
Db 103 CCTCGCCCGCGGACACACGCTCTCTGTA 74
|||||

```

```

RESULT 2
US-09-734-675-3
; Sequence 3, Application US/09734675
; Patent No. 6365391
; GENERAL INFORMATION:
; APPLICANT: WEBSTER, Marion et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND

```

```

; TITLE OF INVENTION:  USES THEREOF
; FILE REFERENCE:  CLO00862
; CURRENT APPLICATION NUMBER:  US/09/734,675
; CURRENT FILING DATE:  2000-12-13
; NUMBER OF SEQ ID NOS:  4
; SOFTWARE:  FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH:  38844
; TYPE:  DNA
; ORGANISM:  Human
US-09-734-675-3

      Query Match      12.4%;  Score 43;  DB 4;  Length 38844;
Best Local Similarity 55.8%;  Pred. No. 0.043;
Matches 82;  Conservative 0;  Mismatches 65;  Indels 0;  Gaps 0;

QY  201  TTTTTCATTCACAGTTCGAGATTGTGGAGTGTTTTTACATTCATCTTTGCTAATGC 260
Db  24453  TTCTTTAAAAGCAATTTTACTNTTAGGAAGTGGTTTAAAGCAATTTTAAATATCA 24512

QY  261  AGTTAGCAGTATGTTTTCATGATGATGATTAATAATCCTTGAATCATATAAAAAAAA 320
Db  24513  TTTATGCAAGAGTTTTCAGGTTTTCATCTTAAACCTTTAAACCAAAAAAAA 24572

QY  321  AAAAATGCTTTGGAACTTGAAAAAA 347
Db  24573  AAAGATTATGTGAATTCGAAGTAAA 24599

RESULT 3
US-07-867-106-2/c
; Sequence 2, Application US/07867106
; Patent No. 5389526
; GENERAL INFORMATION:
; APPLICANT:  Slade, Martin B
; APPLICANT:  Chang, Andy C M
; APPLICANT:  Williams, Keith L
; TITLE OF INVENTION:  Improved Plasmid Vectors for Cellular
; TITLE OF INVENTION:  Sline Moulds of the Genus Dictyostelium
; NUMBER OF SEQUENCES:  19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris
; STREET:  One Liberty Place 46th Floor
; CITY:  Philadelphia
; STATE:  PA
; COUNTRY:  USA
; ZIP:  19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE:  Floppy disk
; COMPUTER:  IBM PC compatible
; OPERATING SYSTEM:  PC-DOS/MS-DOS
; SOFTWARE:  Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER:  US/07/867,106
; FILING DATE:  19920625
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:  AU PJ 7187
; APPLICATION NUMBER:  PCT/AU90/00530
; FILING DATE:  02-NOV-1989
; ATTORNEY/AGENT INFORMATION:
; NAME:  Peeney, Joanne Longo
; REGISTRATION NUMBER:  35,134
; REFERENCE/DOCKET NUMBER:  RICE-0002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE:  215-568-3100
; TELEFAX:  215-568-3439
; INFORMATION FOR SEQ ID NO:  2:
; SEQUENCE CHARACTERISTICS:
; LENGTH:  5852 base pairs
; TYPE:  NUCLEIC ACID
; STRANDEDNESS:  single
; TOPOLOGY:  linear
; MOLECULE TYPE:  DNA (genomic)

```

```

; ANTI-SENSE:  NO
; FEATURE:
; NAME/KEY:  CDS
; LOCATION:  2378..5038
; FEATURE:
; NAME/KEY:  CDS
; LOCATION:  2378..5038
US-07-867-106-2

      Query Match      11.7%;  Score 40.8;  DB 1;  Length 5852;
Best Local Similarity 54.7%;  Pred. No. 0.094;
Matches 81;  Conservative 0;  Mismatches 67;  Indels 0;  Gaps 0;

QY  201  TTTTTCATTCACAGTTCGAGATTGTGGAGTGTTTTTACATTCATCTTTGCTAATGC 260
Db  5767  TTTAATTTTGTATTTTGTATTTTATATATATGTTATTTGTTGTTTACIT 5708

QY  261  AGTTAGCAGTATGTTTTCATGATGATGATTAATAATCCTTGAATCATATAAAAAAAA 320
Db  5707  ATATTTCATTTTATTATTAATAATTAATAATTTAAATTTAAATAATAAAAAAAA 5648

QY  321  AAAAATGCTTTGGAACTTGAAAAAA 348
Db  5647  AAAAAAATAATTTAAATTTAAAAAA 5620

RESULT 4
US-08-345-756-8/c
; Sequence 8, Application US/08345756
; Patent No. 5633438
; GENERAL INFORMATION:
; APPLICANT:  Baszczyński, Chris
; APPLICANT:  Barbour, Eric
; APPLICANT:  Hattori, Jiro
; APPLICANT:  Miki, Brian
; TITLE OF INVENTION:  MICROSPORE-SPECIFIC REGULATORY ELEMENT
; NUMBER OF SEQUENCES:  8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE:  Foley & Lardner
; STREET:  3000 K Street, N.W., Suite 500
; CITY:  Washington
; STATE:  D.C.
; COUNTRY:  USA
; ZIP:  20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE:  Floppy disk
; COMPUTER:  IBM PC compatible
; OPERATING SYSTEM:  PC-DOS/MS-DOS
; SOFTWARE:  Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER:  US/08/345,756
; FILING DATE:  22-NOV-1994
; CLASSIFICATION:  800
; ATTORNEY/AGENT INFORMATION:
; NAME:  BENT, STEPHEN A.
; REGISTRATION NUMBER:  29,768
; REFERENCE/DOCKET NUMBER:  33229/236/PIHI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE:  (202)672-5300
; TELEFAX:  (202)672-5399
; TELEX:  904136
; INFORMATION FOR SEQ ID NO:  8:
; SEQUENCE CHARACTERISTICS:
; LENGTH:  859 base pairs
; TYPE:  nucleic acid
; STRANDEDNESS:  single
; TOPOLOGY:  linear
; MOLECULE TYPE:  DNA (genomic)
US-08-345-756-8

      Query Match      11.3%;  Score 39.2;  DB 1;  Length 859;
Best Local Similarity 54.1%;  Pred. No. 0.14;
Matches 80;  Conservative 0;  Mismatches 68;  Indels 0;  Gaps 0;

```

QY 197 AAACATTTTTCATTCAGAGTTCAGAAATGTGGAGTGTGTTTACATTCATCTTTTGCTA 256
 |||||
 Db 240 AATTTTGGTTTGAATTTTGTTCGAATCATGTGGGTTTCGATTCAGATTCGAGTTCGGGTA 181
 |||||
 QY 257 ATGCAGTTAGCAGTATGTTTTCGATGATGATTAATAATCCTTGAATCATATAAAAAA 316
 |||||
 Db 180 CCATTTTAATATGATGATTTTCTTAAAAAAATTCAAATATACTTAAATCGTCAAAATCA 121
 |||||
 QY 317 AAAAAAATGCTTTTGGAACTTGAAAA 344
 |||||
 Db 120 AAAAAATAAATAATAAAACATAAAAA 93
 |||||

RESULT 5
 US-08-625-198-8/c
 ; Sequence 8, Application US/08625198
 ; Patent No. 5756324
 ; GENERAL INFORMATION:
 ; APPLICANT: Baszczynski, Chris
 ; APPLICANT: Barbour, Eric
 ; APPLICANT: Hattori, Jiro
 ; APPLICANT: Mikki, Brian
 ; TITLE OF INVENTION: MICROSPORE-SPECIFIC REGULATORY ELEMENT
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Foley & Lardner
 ; STREET: 3000 K Street, N.W., Suite 500
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20007-5109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/625,198
 ; FILING DATE: 01-APR-1996
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/345,756
 ; FILING DATE: 22-NOV-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BENT, STEPHEN A.
 ; REGISTRATION NUMBER: 29,768
 ; REFERENCE/DOCKET NUMBER: 33229/236/PIHI
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202)672-5300
 ; TELEFAX: (202)672-5399
 ; TELEX: 904136
 ; INFORMATION FOR SEQ ID NO: 8:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 859 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; US-08-625-198-8

Query Match 11.3%; Score 39.2; DB 1; Length 859;
 Best Local Similarity 54.1%; Pred. No. 0.14; Mismatches 0; Indels 0; Gaps 0;
 Matches 80; Conservative 0
 QY 197 AAACATTTTTCATTCAGAGTTCAGAAATGTGGAGTGTGTTTACATTCATCTTTTGCTA 256
 |||||
 Db 240 AATTTTGGTTTGAATTTTGTTCGAATCATGTGGGTTTCGATTCAGATTCGAGTTCGGGTA 181
 |||||
 QY 257 ATGCAGTTAGCAGTATGTTTTCGATGATGATTAATAATCCTTGAATCATATAAAAAA 316
 |||||
 Db 180 CCATTTTAATATGATGATTTTCTTAAAAAAATTCAAATATACTTAAATCGTCAAAATCA 121
 |||||

QY 317 AAAAAAATGCTTTTGGAACTTGAAAA 344
 |||||
 Db 120 AAAAAATAAATAATAAAACATAAAAA 93
 |||||
 RESULT 6
 US-08-345-756-5/c
 ; Sequence 5, Application US/08345756
 ; Patent No. 5633438
 ; GENERAL INFORMATION:
 ; APPLICANT: Baszczynski, Chris
 ; APPLICANT: Barbour, Eric
 ; APPLICANT: Hattori, Jiro
 ; APPLICANT: Mikki, Brian
 ; TITLE OF INVENTION: MICROSPORE-SPECIFIC REGULATORY ELEMENT
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Foley & Lardner
 ; STREET: 3000 K Street, N.W., Suite 500
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20007-5109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/345,756
 ; FILING DATE: 22-NOV-1994
 ; CLASSIFICATION: 800
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BENT, STEPHEN A.
 ; REGISTRATION NUMBER: 29,768
 ; REFERENCE/DOCKET NUMBER: 33229/236/PIHI
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202)672-5300
 ; TELEFAX: (202)672-5399
 ; TELEX: 904136
 ; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2039 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; US-08-345-756-5

Query Match 11.3%; Score 39.2; DB 1; Length 2039;
 Best Local Similarity 54.1%; Pred. No. 0.18; Mismatches 0; Indels 0; Gaps 0;
 Matches 80; Conservative 0
 QY 197 AAACATTTTTCATTCAGAGTTCAGAAATGTGGAGTGTGTTTACATTCATCTTTTGCTA 256
 |||||
 Db 231 AATTTTGGTTTGAATTTTGTTCGAATCATGTGGGTTTCGATTCAGATTCGAGTTCGGGTA 172
 |||||
 QY 257 ATGCAGTTAGCAGTATGTTTTCGATGATGATTAATAATCCTTGAATCATATAAAAAA 316
 |||||
 Db 171 CCATTTTAATATGATGATTTTCTTAAAAAAATTCAAATATACTTAAATCGTCAAAATCA 112
 |||||
 QY 317 AAAAAAATGCTTTTGGAACTTGAAAA 344
 |||||
 Db 111 AAAAAATAAATAATAAAACATAAAAA 84
 |||||

RESULT 7
 US-08-625-198-5/c
 ; Sequence 5, Application US/08625198
 ; Patent No. 5756324
 ; GENERAL INFORMATION:
 ; APPLICANT: Baszczynski, Chris
 ; APPLICANT: Barbour, Eric

APPLICANT: Hattori, Jiro
APPLICANT: Miki, Brian
TITLE OF INVENTION: MICROSPORE-SPECIFIC REGULATORY ELEMENT
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/625,198
FILING DATE: 01-APR-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/345,756
FILING DATE: 22-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: BENT, STEPHEN A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 33229/236/PIHI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO.: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2039 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-625-198-5

Query Match 11.3%; Score 39.2; DB 1; Length 2039;
Best Local Similarity 54.1%; Pred. No. 0.18;
Matches 80; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 197 AAACATTTTCAATTGCAGTGTCGAGATTGGCGAGTTTTTACATTGTCTTTTGCTA 256
DB 231 AATTITGGTTGAATTTTTGTTCGAATCATTTGCCGTTCCAGATTCGAGTTCGGTA 172
QY 257 ATGCAGTAGCAGTAGTGTTCGACTGTATGACTTAATAAAATCCCTTGATCATAAAAAAA 316
DB 171 CCCATTTTAAATATGCTATTTTTTCTTAAAAAAATCCAATATACATAATCGTCAAATCA 112
QY 317 AAAAAAAAAATGCTTTTGGAACTTGAAAA 344
DB 111 AAAAATAAAATAATAATAAACATAAAAA 84

RESULT 8
US-09-205-258-232
Sequence 232, Application US/09205258
Patent No. 6525174
GENERAL INFORMATION:
APPLICANT: Young et al.
TITLE OF INVENTION: 207 Human Secreted Proteins
FILE REFERENCE: PZ007P1
CURRENT APPLICATION NUMBER: US/09/205,258
CURRENT FILING DATE: 1998-12-04
REGISTERED CLAIMS: PCT/US98/11422
EARLIER APPLICATION NUMBER: 60/048,895
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,895
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,895
EARLIER FILING DATE: 1997-06-06

EARLIER APPLICATION NUMBER: 60/048,881
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,880
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,896
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049,020
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,876
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,895
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,884
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,894
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,971
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,964
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,882
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,899
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,893
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,900
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,901
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049,019
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,970
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,972
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,916
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049,373
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,875
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049,374
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,917
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,949
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,974
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,883
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,897
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,898
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,962
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,963
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,877
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,878
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/070,923
EARLIER FILING DATE: 1997-12-18
EARLIER APPLICATION NUMBER: 60/092,921
EARLIER FILING DATE: 1998-07-15
EARLIER APPLICATION NUMBER: 60/094,657

```

; EARLIER FILING DATE: 1998-07-30
; NUMBER OF SEQ ID NOS: 1227
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 232
; LENGTH: 760
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (438)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (741)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-205-258-232

```

	Query Match	11.1%;	Score 38.8;	DB 4;	Length 760;
	Best Local Similarity	55.1%;	Pred. No. 0.1%;		
	Matches	76;	Conservative	0;	Mismatches 62; Indels 0; Gaps 0;
QY	211	TGACAGTGCAGAAATCTGCAGTGTCTTTTACATGATCTTTGGCTAATGCGAGTATGACGAT	270		
Db	602	TGAAGCTATGTACTAATCTTATGTTTGTAAATGTCTTTTGCATATGCATCATCTTCTTT	661		
QY	271	ATCTTTTGGCATGTATGACTTAAATAAATCCCTTGAATCATATAAAAAAAAAAAAAAAAAATGCT	330		
Db	662	TCCTTTTGTAGGTATAAATAAACAACACTGTTGCATATAAAAAAAAAAAAAAAAAAAAAA	721		
QY	331	TTGGAACTTGAAAAAAA	348		
Db	722	AAAAAAAAAAAAAAAAAAAA	739		

```

RESULT 9
US-09-205-258-103
/ Sequence 103, Application US/09205258
/ Patent No. 6525174
/ GENERAL INFORMATION:
/ APPLICANT: Young et al.
/ TITLE OF INVENTION: 207 Human Secreted Proteins
/ FILE REFERENCE: PZ007F1
/ CURRENT APPLICATION NUMBER: US/09/205,258
/ CURRENT FILING DATE: 1998-12-04
/ EARLIER APPLICATION NUMBER: PCT/US98/11422
/ EARLIER FILING DATE: 1998-06-04
/ EARLIER APPLICATION NUMBER: 60/048,885
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/049,375
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,881
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,880
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,896
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/049,020
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,876
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,895
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,884
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,894
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,971
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,964
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,892
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/048,899

```

```

? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,893
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,900
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,901
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,892
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,915
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/049,019
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,970
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,972
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,916
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/049,373
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,875
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/049,374
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,917
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,949
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,974
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,883
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,897
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,898
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,962
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,963
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/048,877
? EARLIER FILING DATE: 1997-06-06
? EARLIER APPLICATION NUMBER: 60/070,923
? EARLIER FILING DATE: 1997-12-18
? EARLIER APPLICATION NUMBER: 60/092,921
? EARLIER FILING DATE: 1998-07-15
? EARLIER APPLICATION NUMBER: 60/094,657
? EARLIER FILING DATE: 1998-07-30
? NUMBER OF SEQ ID NOS: 1227
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 103
? LENGTH: 2218
? TYPE: DNA
? ORGANISM: Homo sapiens
US-09-0205-258-103

```

	Query Match	11.1%	Score 38.8	DB 4	Length 2218	
	Best Local Similarity	55.1%	Pred. No. 0.24			
	Matches 76	Conservative 0	Mismatches 62	Indels 0	Gaps 0	
QY	211	TGACAGTTGCAGAAATGTGGAGTGTCTTTTACATTGATCTTTTTCGTAAATGCAGTTAGCAGT	270			
Db	2061	TGAAGAGTATGTTACTATTCTTAGTCTTGTAATGTCCCTTTTCATACCATCATCTTCGTTT	2120			
QY	271	ATGTTTTCGATGTATGACTTAATAAATCCCTTGAATCATATAAAAAAATAAAAAATGCT	330			
Db	2121	TCCTTTTGTAGGTATAATAAAAAACACTGTGACATAAAAAAATAAAAAAATAAAAA	2180			
QY	331	TTGGAACTTGAAAAAAA	348			

DB 2181 AAAAAAAAAAAAAAAAAA 2198

RESULT 10

US-09-277-716-17

Sequence 17, Application US/09277716A

Patent No. 6232107

GENERAL INFORMATION:

APPLICANT: Bryan, Bruce

APPLICANT: Szent-Gyorgyi, Christopher

APPLICANT: PROMUNE, LTD.

TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE

CURRENT FILING DATE: 1999-03-26

EARLIER APPLICATION NUMBER: 60/102,939

EARLIER FILING DATE: 1998-10-01

EARLIER APPLICATION NUMBER: 60/089,367

EARLIER FILING DATE: 1998-06-15

EARLIER APPLICATION NUMBER: 60/079,624

EARLIER FILING DATE: 1998-03-27

NUMBER OF SEQ ID NOS: 32

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 17

LENGTH: 1217

TYPE: DNA

ORGANISM: Renilla mulleri

FEATURE:

NAME/KEY: CDS

LOCATION: (31)..(963)

OTHER INFORMATION: Renilla mulleri luciferase

US-09-277-716-17

Query Match 10.6%; Score 36.8; DB 3; Length 1217;

Best Local Similarity 63.6%; Pred. No. 0.69;

Matches 56; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 238 TTACATGATCTTTCTTAATGACGTAGCAGTATGTTGCATGTATGACTTAATAAT 297

Db 1126 TTCTCTTAGCTATTATATAATACATCTTATGTAATAAAACCTTTGTTTATAAAT 1185

QY 298 CCTTGAATCATAAAAAAAAAAAAAA 325

Db 1186 TAATGATTCAGAAAAAAAAAAAAA 1213

RESULT 12

US-08-745-995A-31

Sequence 31, Application US/08745995A

Patent No. 6372887

GENERAL INFORMATION:

APPLICANT: Silver, Gary M.

APPLICANT: Wisniewski, Nancy

TITLE OF INVENTION: No. 6372887el Serine Protease Inhibitor

TITLE OF INVENTION: Nucleic Acid Molecules, Proteins

TITLE OF INVENTION: and Uses Thereof

NUMBER OF SEQUENCES: 41

CORRESPONDENCE ADDRESS:

ADDRESSEE: Heska Corporation

STREET: 1825 Sharp Point Drive

CITY: Fort Collins

STATE: Colorado

COUNTRY: USA

ZIP: 80525

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: Windows 95

SOFTWARE: WordPerfect for Windows, Version 7.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/745,995A

FILING DATE: 07-NOV-1996

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Verser, Carol Talkington

REGISTRATION NUMBER: 37,459

REFERENCE/DOCKET NUMBER: FC-2

TELECOMMUNICATION INFORMATION:

TELEPHONE: 970/493-7272

TELEFAX: 970/484-9505

INFORMATION FOR SEQ ID NO: 31:

SEQUENCE CHARACTERISTICS:

LENGTH: 1454 nucleotides

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 20..1210

US-08-745-995A-31

Query Match 10.5%; Score 36.4; DB 4; Length 1454;

Best Local Similarity 58.2%; Pred. No. 0.93;

Matches 64; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 216 GTTCAGAAATGTCAGCTGTTTACATTCATCTTTGCTAAATGACGTATGACTTAT 275

Db 1345 GTAGACGAAATGTTTGTGTTTGTAGTTTTCATCTTTTATGTAATGTAATACCTATAAT 1404

DB 2181 AAAAAAAAAAAAAAAAAA 2198

RESULT 10

US-09-277-716-17

Sequence 17, Application US/09277716A

Patent No. 6232107

GENERAL INFORMATION:

APPLICANT: Bryan, Bruce

APPLICANT: Szent-Gyorgyi, Christopher

APPLICANT: PROMUNE, LTD.

TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE

CURRENT FILING DATE: 1999-03-26

EARLIER APPLICATION NUMBER: 60/102,939

EARLIER FILING DATE: 1998-10-01

EARLIER APPLICATION NUMBER: 60/089,367

EARLIER FILING DATE: 1998-06-15

EARLIER APPLICATION NUMBER: 60/079,624

EARLIER FILING DATE: 1998-03-27

NUMBER OF SEQ ID NOS: 32

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 17

LENGTH: 1217

TYPE: DNA

ORGANISM: Renilla mulleri

FEATURE:

NAME/KEY: CDS

LOCATION: (31)..(963)

OTHER INFORMATION: Renilla mulleri luciferase

US-09-277-716-17

Query Match 10.6%; Score 36.8; DB 3; Length 1217;

Best Local Similarity 63.6%; Pred. No. 0.69;

Matches 56; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 238 TTACATGATCTTTCTTAATGACGTAGCAGTATGTTGCATGTATGACTTAATAAT 297

Db 1126 TTCTCTTAGCTATTATATAATACATCTTATGTAATAAAACCTTTGTTTATAAAT 1185

QY 298 CCTTGAATCATAAAAAAAAAAAAAA 325

Db 1186 TAATGATTCAGAAAAAAAAAAAAA 1213

RESULT 11

US-09-609-161B-17

Sequence 17, Application US/09609161B

Patent No. 6436682

GENERAL INFORMATION:

APPLICANT: Bryan, Bruce

APPLICANT: Szent-Gyorgyi, Christopher

APPLICANT: PROMUNE, LTD.

TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE LUC

TITLE OF INVENTION: AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HIG

TITLE OF INVENTION: SCREENING AND NOVELTY ITEMS

FILE REFERENCE: 24729-121B

CURRENT APPLICATION NUMBER: US/09/609,161B

CURRENT FILING DATE: 2000-06-30

PRIOR APPLICATION NUMBER: 09/277,716

PRIOR FILING DATE: 1999-03-26

PRIOR APPLICATION NUMBER: 60/102,939

PRIOR FILING DATE: 1998-10-01

PRIOR APPLICATION NUMBER: 60/089,367

PRIOR FILING DATE: 1998-06-15

PRIOR APPLICATION NUMBER: 60/079,624

PRIOR FILING DATE: 1998-03-27

NUMBER OF SEQ ID NOS: 32

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 17

LENGTH: 1217

TYPE: DNA

QY 276 TTGCATGTATGACTTAATAAATCCTTGAAATCATAAAAAAAAAAAAAAAAA 325
 Db 1405 GTTGAGTTTATGTATATAAATGTTAAATGTGAAAAAAAAAAAAAAAA 1454

RESULT 13

US-08-745-995A-33/c
 ; Sequence 33, Application US/08745995A
 ; Patent No. 6372887

; GENERAL INFORMATION:

; APPLICANT: Silver, Gary M.
 ; APPLICANT: Wisniewski, Nancy
 ; TITLE OF INVENTION: No. 6372887el Serine Protease Inhibitor
 ; TITLE OF INVENTION: Nucleic Acid Molecules, Proteins
 ; TITLE OF INVENTION: and Uses Thereof
 ; NUMBER OF SEQUENCES: 41

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Heska Corporation
 ; STREET: 1825 Sharp Point Drive
 ; CITY: Fort Collins
 ; STATE: Colorado
 ; COUNTRY: USA
 ; ZIP: 80525

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: Windows 95
 ; SOFTWARE: WordPerfect for Windows, Version 7.0
 ; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/745.995A

; FILING DATE: 07-NOV-1996

; CLASSIFICATION: 536

; ATTORNEY/AGENT INFORMATION:

; NAME: Verser, Carol Talkington

; REGISTRATION NUMBER: 37,459

; REFERENCE/DOCKET NUMBER: FC-2

; TELEPHONE: 970/493-7272

; TELEFAX: 970/484-9505

; INFORMATION FOR SEQ ID NO: 33:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1454 nucleotides

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

US-08-745-995A-33

Query Match 10.5%; Score 36.4; DB 4; Length 1454;
 Best Local Similarity 58.3%; Pred. No. 0.93; Indels 0; Gaps 0;
 Matches 64; Conservative 0; Mismatches 0

QY 216 GTTCAGAAATGTGGAGTGTCTTTTACATGCTTTTGCTTAATGCAGTTAGCAGTATGTT 275
 Db 110 GTAGACGAAATGTTTGTGTTTGTAGTTTTCACCTTTTATGAATGTAATCACCCTATATAAT 51

QY 276 TTGCATGTATGACTTAATAAATCCTTGAAATCATAAAAAAAAAAAAAAAAA 325
 Db 50 GTTGATGTTTATGTATAAATGTTAAATGTGAAAAAAAAAAAAAAAAAAAA 1

RESULT 14

US-09-005-352-31
 ; Sequence 31, Application US/09005352
 ; Patent No. 6479253

; GENERAL INFORMATION:

; APPLICANT: Silver, Gary M.
 ; APPLICANT: Wisniewski, Nancy
 ; TITLE OF INVENTION: No. 6479253el Serine Protease Inhibitor
 ; TITLE OF INVENTION: Nucleic Acid Molecules, Proteins
 ; TITLE OF INVENTION: and Uses Thereof
 ; NUMBER OF SEQUENCES: 41

; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Heska Corporation
 ; STREET: 1825 Sharp Point Drive
 ; CITY: Fort Collins
 ; STATE: Colorado
 ; COUNTRY: USA
 ; ZIP: 80525
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: Windows 95
 ; SOFTWARE: WordPerfect for Windows, Version 7.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/005.352
 ; FILING DATE: 07-NOV-1996
 ; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Verser, Carol Talkington
 ; REGISTRATION NUMBER: 37,459
 ; REFERENCE/DOCKET NUMBER: FC-2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 970/493-7272
 ; TELEFAX: 970/484-9505

; INFORMATION FOR SEQ ID NO: 31:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1454 nucleotides

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 20..1210

US-09-005-352-31

Query Match 10.5%; Score 36.4; DB 4; Length 1454;
 Best Local Similarity 58.2%; Pred. No. 0.93;
 Matches 64; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 216 GTTCAGAAATGTGGAGTGTCTTTTACATGCTTTTGCTTAATGCAGTTAGCAGTATGTT 275
 Db 1345 GTAGACGAAATGTTTGTGTTTGTAGTTTTCACCTTTTATGAATGTAATCACCCTATATAAT 1404
 QY 276 TTGCATGTATGACTTAATAAATCCTTGAAATCATAAAAAAAAAAAAAAAAA 325
 Db 1405 GTTGATGTTTATGTATAAATGTTAAATGTGAAAAAAAAAAAAAAAAAAAA 1454

RESULT 15

US-09-005-352-33/c
 ; Sequence 33, Application US/09005352
 ; Patent No. 6479253

; GENERAL INFORMATION:

; APPLICANT: Silver, Gary M.
 ; APPLICANT: Wisniewski, Nancy
 ; TITLE OF INVENTION: No. 6479253el Serine Protease Inhibitor
 ; TITLE OF INVENTION: Nucleic Acid Molecules, Proteins
 ; TITLE OF INVENTION: and Uses Thereof
 ; NUMBER OF SEQUENCES: 41
 ; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Heska Corporation
 ; STREET: 1825 Sharp Point Drive
 ; CITY: Fort Collins
 ; STATE: Colorado
 ; COUNTRY: USA
 ; ZIP: 80525

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: Windows 95
 ; SOFTWARE: WordPerfect for Windows, Version 7.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/005.352

Mon Nov 24 15:20:12 2003

```

;
; FILING DATE: 07-NOV-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Verser, Carol Talkington
; REGISTRATION NUMBER: 37,459
; REFERENCE/DOCKET NUMBER: PC-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 970/493-7272
; TELEFAX: 970/484-9505
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1454 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
;
US-09-005-352-33

Query Match          10.5%; Score 36.4; DB 4; Length 1454;
Best Local Similarity 58.2%; Pred. No. 0.93; 46; Indels 0; Gaps 0;
Matches 64; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY      216 GTTCGAGATTGTGGAGTGTGTTTACATTGATCTTTTGCTAATGCAGTTAGCAGTAATGTT 275
Db      110 GTAGACGAAAAGTTTGTGTTTGTAGTTTTCACCTTTTATGAATGTAATCACCTATATAAT 51

QY      276 TTGCATGTAAGACTTAATAAATCCTTGAATCATATAAAAAAAAAAAAAAAAA 325
Db      50 GTTGTAGTTTATGTAATAAATAATGTTAAATGTGAAAAAAAAAAAAAAAAAAAA 1

Search completed: November 23, 2003, 15:53:46
Job time : 28.5359 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 23, 2003, 14:02:08 : Search time 123.329 Seconds
(without alignments)
9221.962 Million cell updates/sec

Title: US-09-717-321A-17

Perfect score: 348

Sequence: 1 tgaacatcactgtctgcc.....ctttggaacttgaaaaaaa 348

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2169961 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	211.6	60.8	2941	14	US-10-198-846-9874
2	209.4	60.2	2051	9	US-09-925-302-255
3	192.4	55.3	1644	14	US-10-153-668-48
4	192.4	55.3	1646	14	US-10-153-668-46
5	187.4	53.9	1232	10	US-09-880-107-1632
6	163.6	47.0	353	10	US-09-960-352-11432
7	116.4	33.4	3740	11	US-09-764-891-9984
8	95.4	27.4	201	10	US-09-960-352-4357
9	52.4	15.1	162	10	US-09-920-300A-1549
10	52.4	15.1	162	12	US-10-099-926-1549
11	52.4	15.1	162	13	US-10-033-528-1549
12	52.4	15.1	477	14	US-10-066-543-3022
13	52.4	15.1	549	14	US-10-066-543-2878
14	52.4	15.1	567	14	US-10-066-543-2364
15	52.4	15.1	594	14	US-10-066-543-2864
16	52.4	15.1	605	11	US-09-871-161-483

17	51.8	14.9	2313	14	US-10-198-846-12919	Sequence 12919, A
c 18	50.8	14.6	602	10	US-09-878-178-1654	Sequence 1654, A
c 19	50.8	14.6	602	13	US-10-046-935-1654	Sequence 1654, A
c 20	50.8	14.6	602	14	US-10-146-502-1654	Sequence 1654, A
c 21	45	12.9	249	12	US-09-814-353-1384	Sequence 1384, A
c 22	45	12.9	249	12	US-09-814-353-7747	Sequence 7747, A
c 23	43.2	12.4	424	10	US-09-814-353-14132	Sequence 14132, A
c 24	43.2	12.4	424	10	US-09-814-353-14132	Sequence 14132, A
c 25	43	12.4	38844	13	US-10-060-333-3	Sequence 1218, A
c 26	42.2	12.1	640681	10	US-09-790-988-1	Sequence 3, Appli
c 27	42	12.1	15261	11	US-09-764-891-10186	Sequence 1, Appli
c 28	41.8	12.0	9817	12	US-10-311-455-1341	Sequence 10186, A
c 29	41.4	11.9	6464	12	US-10-311-455-1341	Sequence 1341, A
c 30	40.6	11.7	336	12	US-09-814-353-4742	Sequence 487, App
c 31	40.6	11.7	336	12	US-09-814-353-4742	Sequence 4742, App
c 32	40.4	11.6	469	11	US-09-918-995-13017	Sequence 11039, A
c 33	39.8	11.4	3673778	12	US-10-312-841-1	Sequence 13017, A
c 34	39.4	11.3	6059	12	US-10-311-455-786	Sequence 1, Appli
c 35	39.2	11.3	277	10	US-09-960-352-12673	Sequence 786, App
c 36	39.2	11.3	375	10	US-09-960-352-12673	Sequence 12673, A
c 37	39.2	11.3	6621	12	US-10-311-455-886	Sequence 15014, A
c 38	39	11.2	501	12	US-10-027-632-184148	Sequence 886, App
c 39	39	11.2	501	13	US-10-027-632-184148	Sequence 184148, A
c 40	39	11.2	1460	14	US-10-198-846-9893	Sequence 184148, A
c 41	38.8	11.1	760	12	US-09-933-767-103	Sequence 9893, App
c 42	38.8	11.1	760	14	US-10-023-282-232	Sequence 103, App
c 43	38.8	11.1	2218	12	US-09-933-767-232	Sequence 232, App
c 44	38.8	11.1	2218	14	US-10-023-282-103	Sequence 103, App
c 45	38.8	11.1	2261	11	US-09-764-891-10185	Sequence 10185, A

ALIGNMENTS

RESULT 1

US-10-198-846-9874
; Sequence 9874, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9874
; LENGTH: 2941
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-198-846-9874

Query Match 60.8%; Score 211.6; DB 14; Length 2941;
Best Local Similarity 80.2%; Pred. No. 1.6e-45;
Matches 295; Conservative 0; Mismatches 49; Indels 24; Gaps 3;

QY 4 AACATCAGTCTTGGCAGGTACCAACACTGTCATGTCATGTCGCGCCCTCTAGA 63
|||||
Db 2106 AACATCAGTCTTGGCAGGTACCAACACTGTCATGTCATGTCGCGCCCTCTAGA 2164
|||||
QY 64 CTCACCCACGGGACACATGCTCCGTA-----CCTTTGGGT 102
|||||
Db 2165 CCTCGCCACGGGACACACAGCGCTCCTGTAGTCGTTTCCTTATGATGTCCTTTGGT 2224
|||||
QY 103 CTGTAGGTTCTGTCAA--GCCTAGTGTACGGCGTCTGTACAACTCACTACTG 160
|||||

```
Db 2225 CTGTGAGGTTCTGTAACTGTGCTAGTGTGACGAGTGTCTGTACAACCTTAACCTCACTGG 2284
Qy 161 CAAGAACACAGTGTGGGCTTTTCGACCACCTAGAACAACTTTTTCACAGTTGC 220
Db 2285 CGAGNATACAGGTGGGACCCCTTCAGCCACTACACAGAAATTTTAAATTCACAGTTGC 1904
Qy 221 AGAATTGTGGAGTGTGTTTACATTGATCTTTTGTAAATGACAGTATGAGCAGTATGTTTGA 280
Db 2345 AGAATTGTGGAGTGTGTTTACATTGATCTTTTGTAAATGACAGTATGAGCAGTATGTTTGA 1964
Qy 281 TGTATGACTTAATAATCCTTGAATCATAAAAAAATAAAAAA 325
Db 1965 TGTATGACTTAATAATCCTTGAATCATAAAAAAATAAAAAA 2009

RESULT 3
US-10-153-668-48
; Sequence 48, Application US/10153668
; Publication No. US20030092616A1
; GENERAL INFORMATION:
; APPLICANT: HONDA, Goichi
; APPLICANT: MATSUDA, Akio
; APPLICANT: MURAMATSU, Shuji
; APPLICANT: ISHIZAWA, Kenya
; TITLE OF INVENTION: STAT6 Activating Gene
; FILE REFERENCE: 1254-0207P
; CURRENT APPLICATION NUMBER: US/10/153,668
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: US 60/293,172
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/316,031
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/328,403
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: JP 2001-157043
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: JP 2001-260681
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: JP 2001-313175
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 488
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 48
; LENGTH: 1644
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (109)..(447)
; US-10-153-668-48

Query Match 55.3%; Score 192.4; DB 14; Length 1644;
Best Local Similarity 81.7%; Pred. No. 1.4e-40;
Matches 288; Conservative 0; Mismatches 36; Indels 24; Gaps 3;

Qy 4 AACATCACTGTCTTGGCAGAGTACCAACTGTCATGTGATGATGCGGCCCTCTAGA 63
Db 1318 AACATCACTGTCTTGGCAGAGTACCAACTGTCATGTGATGATGCGGCCCTCTAGA 1376

Qy 64 CCTCACCACGCGGACACATGCTCCGGTA-----CCTTTGGGT 102
Db 1377 CCTCGCCACGCGGACACACGCTCTCTGTAGTGCCTTGTGATGTTGCTTTGGGT 1436

Qy 103 CTGTGAGGTCTCTGCTCAA--GCGCTAGTGTCTAAAGCCGCTTCTGTACAACTTAACCTCTGG 160
Db 1437 CTGTGAGGTCTCTGCTCAA--GCGCTAGTGTCTAAAGCCGCTTCTGTACAACTTAACCTCTGG 1496

Qy 161 CAAGAACACAGTGTGGGCTTTTCGACCACCTAGAACAACTTTTTCACAGTTGC 220
Db 1497 CGAGNATACAGGTGGGACCCCTTCAGCCACTACACAGAAATTTTAAATTCACAGTTGC 1556

Qy 221 AGAATTGTGGAGTGTGTTTACATTGATCTTTTGTAAATGACAGTATGAGCAGTATGTTTGA 280
Db 1557 AGAATTGTGGAGTGTGTTTACATTGATCTTTTGTAAATGACAGTATGAGCAGTATGTTTGA 1616
```

```
Db 2225 CTGTGAGGTTCTGTAACTGTGCTAGTGTGACGAGTGTCTGTACAACCTTAACCTCACTGG 2284
Qy 161 CAAGAACACAGTGTGGGCTTTTCGACCACCTAGAACAACTTTTTCACAGTTGC 220
Db 2285 CGAGNATACAGGTGGGACCCCTTCAGCCACTACACAGAAATTTTAAATTCACAGTTGC 2344
Qy 221 AGAATTGTGGAGTGTGTTTACATTGATCTTTTGTAAATGACAGTATGAGCAGTATGTTTGA 280
Db 2345 AGAATTGTGGAGTGTGTTTACATTGATCTTTTGTAAATGACAGTATGAGCAGTATGTTTGA 2404
Qy 281 TGTATGACTTAATAATCCTTGAATCATAAAAAAATAAAAAA 340
Db 2405 TGTATGACTTAATAATCCTTGAATCATAAAAAAATAAAAAA 2464

Qy 341 AAAAAAA 348
Db 2465 AAAAAAA 2472

RESULT 2
US-09-925-302-255
; Sequence 255, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 255
; LENGTH: 2051
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (50)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (68)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (2027)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (2046)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-925-302-255

Query Match 60.2%; Score 209.4; DB 9; Length 2051;
Best Local Similarity 82.6%; Pred. No. 5.2e-45;
Matches 285; Conservative 0; Mismatches 36; Indels 24; Gaps 3;

Qy 4 AACATCACTGTCTTGGCAGAGTACCAACTGTCATGTGATGATGCGGCCCTCTAGA 63
Db 1666 AACATCACTGTCTTGGCAGAGTACCAACTGTCATGTGATGATGCGGCCCTCTAGA 1724

Qy 64 CCTCACCACGCGGACACATGCTCCGGTA-----CCTTTGGGT 102
Db 1725 CCTCGCCACGCGGACACACGCTCTCTGTAGTGCCTTGTGATGTTGCTTTGGGT 1784

Qy 103 CTGTGAGGTCTCTGCTCAA--GCGCTAGTGTCTAAAGCCGCTTCTGTACAACTTAACCTCTGG 160
Db 1785 CTGTGAGGTCTCTGCTCAA--GCGCTAGTGTCTAAAGCCGCTTCTGTACAACTTAACCTCTGG 1844
```

Qy 281 TGTATGACTTAATAAATCCCTGAATCAT 308
|||
Db 1617 TGTATGACTTAATAAATCCCTGAATCAT 1644

```

RESULT 4
US/10-153-668-46
; Sequence 46, Application US/10153668
; Publication No. US2003092616A1
; GENERAL INFORMATION:
; APPLICANT: HONDA, Goichi
; APPLICANT: MATSUDA, Akio
; APPLICANT: MURAMATSU, Shuji
; APPLICANT: ISHIZAWA, Kenya
; TITLE OF INVENTION: STAT6 Activating Gene
; FILE REFERENCE: 1254-0207P
; CURRENT APPLICATION NUMBER: US/10/153,668
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: US 60/293,172
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/316,031
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/328,403
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: JP 2001-157043
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: JP 2001-260681
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: JP 2001-313175
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 488
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46

```

```

; CDNA ID: NC_0148
; LENGTH: 1646
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (109)..(273)
US-10-153-668-46

```

```

Query Match          55.3%; Score 192.4; DB 14; Length 1646;
Best Local Similarity 81.7%; Pred. No. 1,4e-40;
Matches 268; Conservative 0; Mismatches 36; Indels 24; Gaps 3;

      QY      4 AACATCAGCTGTCTTGCCAGAGTACCAACTGTCATGTGTGATGGCGGCCCGCTCTAGA 63
               |||||
      Ddb     1320 AACATCAGCTGTCTTGCCAGATTACCGACACTGTCACCTTGACCAATACTGA-CCTCTTTA 1378

      QY      64 CCTCACCCACGGGCACACATGCTTCGGTA-----CCTTTGGGT 102
               |||||
      Ddb     1379 C CTCGCCACGGGCACACAGCCTCTGTAGTCGCTTTGCTATGTGATGTCCTTTGGGT 1438

```

103	CTGTGAGGTTCTGTCAA--GCGCTAGTGCTAACGCCGCTTCTGTACACCACTAACTCACTGG	160
	QY	
1439	CTGTGAGGTTCTGTAAACTGTGTAGTGTCTGACGATGTTCTGTACAACTTAACTCACTGG	1498
	Db	
161	CAAGAAACACAGTGTGGGCGCTTCGACCACTAGAACAAACTTTTTTCAATTGACAGTTGC	220
	QY	
1499	CGAGAAATACAGCGTGGGACCTTCACCCACTACAACAGAAATTTTTTAAATTGACAGTTGC	1558
	Db	
221	AGAATTGTGGAGTGTTTTACATTGATCTTTTGCTAATGCAGTTAGCAGTATCTTTTGCA	280
	QY	
1559	AGAAATTGTGGAGTGTTTTACATTGATCTTTTGCTAATGCAAATTAGCAATTAGTTTGCA	1618
	Db	
281	TGTATGACTTTAATAATCCTTGAATCAT	308
	QY	
1619	TGTATGACTTTAATAATCCTTGAATCAT	1646
	Db	

RESULT 5
US-09-880-107-1632
; Sequence 1632, Application US/09880107

```

; Patent No. US20020142981A1
;
; GENERAL INFORMATION:
;
;   APPLICANT: Hoine, Darci T.
;   APPLICANT: Vockley, Joseph G.
;   APPLICANT: Scherf, Uwe
;   APPLICANT: Gene Logic, Inc.
;
;   TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
;
;   FILE REFERENCE: 44921-5028-WO
;   CURRENT APPLICATION NUMBER: US/09/880,107
;   CURRENT FILING DATE: 2001-06-14
;   PRIOR APPLICATION NUMBER: US 60/211,379
;   PRIOR FILING DATE: 2000-06-14
;   PRIOR APPLICATION NUMBER: US 60/237,054
;   PRIOR FILING DATE: 2000-10-02
;   NUMBER OF SEQ ID NOS: 3950
;   SOFTWARE: PatentIn Ver. 2.1
;   SEQ ID NO 1632
;   LENGTH: 1232
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. US20020142981A1 D25274
; US-09-880-107-1632

```

Query Match	53.9%	Score 187.4	DB 10	Length 1232
Best Local Similarity	80.9%	Pred. No. 2.6e-39		
Matches 263	Conservative 1	Mismatches 37	Indels 24	Gaps 3

QY	4	AACATCACTGCTCTTGCAGAGTACCAACACTGTCATGTGATTGATGCGGCCCTCTTAGA	63
Db	905	AACATCACTGCTCTTGCAGAGTACCGACACTGTCACCTGACCAATACTGA-CCTCTCTTA	963
QY	64	CCTCACCCACGGGACACATGCTTCGGTA-----CCTTCGGT	102
Db	964	CCTCGCCACGGGACACAGCCCTCTGTAGTCGCTTGCCTATTGATGTCCTTCGGT	1023
QY	103	CTGTGAGGTTCTGTCAA--GCGTAGTGCTAACGCCCTCTGTACAACTCACTCACTGG	160
Db	1024	CTGTGAGGTTCTGTAACTGTGTAGTGTGACGATGTTCTGTACAACTTAACTCACTGG	1083
QY	161	CAAGACACAGTGTTGGGCTTTCGACCACACTAGAACAACTTTTCAATTGACAGTTGC	220
Db	1084	CGAAGATACAGCGTGGGACCTTCAGCCACTACACAGAAATTTTAAATTGACAGTTGC	1143
QY	221	AGAATTGTGGAGTGTTTTACATTGATCTTTTCTAATGCAGTTAGCAGATATGTTTNGCA	280
Db	1144	AGAATTGTGGAGTGTTTTACATTGATCTTTTCTAATGCATATGATTTGTTTNGCA	1203
QY	281	TGATGACTTAATAATCTCTGAAT	305
Db	1204	TGTATGACTTAATAATCTCTTGAA	1228

```

RESULT 6
US-09-960-352-11432/c
; Sequence 11432, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND O
; TITLE OF INVENTION: MUSCLE AND FAT DER
; FILE REFERENCE: 46511.006/37-21 (10298)
; CURRENT APPLICATION NUMBER: US/09/960-3
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 45112
; SEQ ID NO 11432
; LENGTH: 353
; TYPE: DNA
; ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 49-LIB305

```

```
US-09-960-352-11432
Query Match      47.0%; Score 163.6; DB 10; Length 353;
Best Local Similarity 90.3%; Pred. No. 2.8e-33;
Matches 186; Conservative 0; Mismatches 19; Indels 1; Gaps 1;

QY 120 GCGTAGTGTACCGCGCTTCTGTACACCTAATCACTGCGCAAGAACACAGTGTGGC 179
Db 246 GTGCTGCTGCTACCAAGTCTGTACAA-CITACCCTGCGCAAGAACACAGTGTGGC 188
QY 180 CTTTCGACACACAGCAACAACTTTTCAATTGACAGTTGCGAATTTGGAGTGTTTT 239
Db 187 CTTTCGACACACAGCAACAACTTTTCAATTGACAGTTGCGAATTTGGAGTGTTTT 128
QY 240 ACATTGATCTTTTGTCTAATGACAGTTAGCAGTAGTGTTTTGCATGTATGACTTAATAATCC 299
Db 127 ACATTGATCTTTTGTCTAATGACAGTTAGCAGTAGTGTTTTGCATGTATGACTTAATAATCC 68
QY 300 TTGAATCATATAAAAAAATAAAAAA 325
Db 67 TTGAATCATATAAAAAAATAAAAAA 42

RESULT 7
US-09-764-891-9984
; Sequence 9984, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 9984
; LENGTH: 3740
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-9984

Query Match      33.4%; Score 116.4; DB 11; Length 3740;
Best Local Similarity 82.0%; Pred. No. 2.4e-20;
Matches 146; Conservative 0; Mismatches 31; Indels 1; Gaps 1;

QY 147 ACCTAATCACTGCGCAAGAACACAGTGTGGCGCTTTTCGACCACTAGAACAACTTTT 206
Db 2667 ACCTTAAGCACTGCGCAAGAACATATAACACTGGGCTTTCACAACTAGAACAA-TATTT 2725
QY 207 CAATTGACAGTTGCGAATTTGGAGTGTTTTACATTTGATCTTTTGTCTAATGAGTTAG 266
Db 2726 AATGACAGCTTCCAGAAATTTGGGGTAATTTTACATTTGATCTTTTGTCTAATGCAATCAG 2785
QY 267 CAGTATGTTTGCATGTATGACTTAATATCTTGAATCCTTGAATCATAAAAAAATAAAAA 324
Db 2786 CAATGTGTTTGCACATGGAATTTAATAATCCTTTAATCATTAATAAAAAATAAAAA 2843

RESULT 8
US-09-960-352-4357/c
; Sequence 4357, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Bvatt, John C.
; APPLICANT: Mathialagan, Naganpan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
```

```
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 4357
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 19-LIB3058-009-Q1-K1-E3
US-09-960-352-4357

Query Match      27.4%; Score 95.4; DB 10; Length 201;
Best Local Similarity 76.5%; Pred. No. 2.1e-15;
Matches 117; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 174 TTGGCCCTTTTCACCACTAGAACAACTTTTCAATTGACAGTTGCGAATTTGGAGT 233
Db 193 TTGAACCTTTTCAAAAATTAGAACATAATTTTAAATTGACAGTTGCGAATTTGGAGT 134
QY 234 GTTTTACATTCATCTTTTGTCTAATGACAGTTAGCAGTAGTGTTTTGCATGTATGACTTAAT 293
Db 133 GATTTTAGATTGATTTTGTGATAAGCTGATTAAGATATGTTTGAAGGTATCCCTTAAT 74
QY 294 AAATCCTTGAATCATAAAAAATAAAAAAATAAAAAA 326
Db 73 AAATTTCTTGACCCCTCAGTATAAAGATAAAAT 41

RESULT 9
US-09-920-300A-1549
; Sequence 1549, Application US/09920300A
; Patent No. US20020136728A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.547
; CURRENT APPLICATION NUMBER: US/09/920,300A
; CURRENT FILING DATE: 2001-07-31
; NUMBER OF SEQ ID NOS: 1789
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1549
; LENGTH: 162
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-920-300A-1549

Query Match      15.1%; Score 52.4; DB 10; Length 162;
Best Local Similarity 81.1%; Pred. No. 0.00041;
Matches 73; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

QY 4 AACATCACTGTCTTTCGACAGTAGCAACACTGTCTGATTTGATCGCGCCCTCTAGA 63
Db 39 AACATCACTGTCTTTCGACAGTAGCAACACTGTCTGATTTGATCGCGCCCTCTAGA 97
QY 64 CCTCACCGCGGACACATGTTCCGGTA 93
Db 98 CCTCGCCGCGGACACACGCTCTCTGTA 127

RESULT 10
US-10-099-926-1549
; Sequence 1549, Application US/10099926
; Publication No. US20030166064A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; APPLICANT: Jiang, Yugu
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.547C2
```

```

; CURRENT APPLICATION NUMBER: US/10/099,926
; CURRENT FILING DATE: 2002-03-17
; NUMBER OF SEQ ID NOS: 1982
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1549
; LENGTH: 162
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-099-926-1549

Query Match
Best Local Similarity 15.1%; Score 52.4; DB 12; Length 162;
Matches 73; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

QY 4 AACATCACTGTCTTCCAGAGTACCAACACTGTCTGATGATGCGGCCCTCTTAGA 63
Db 39 AACATCACTGTCTTCCAGATTACCGACACTGTCTGACCAATACTGA-CCCTCTTTA 97
QY 64 CCTCACCACCGGGACACATGCTTCCGGTA 93
Db 98 CCTCGCCACCGGGACACACGGCTCTCTGTA 127

RESULT 11
US-10-033-528-1549
; Sequence 1549, Application US/10033528
; Publication No. US20020131971A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 21021.547C1
; CURRENT APPLICATION NUMBER: US/10/033.528
; CURRENT FILING DATE: 2001-12-26
; NUMBER OF SEQ ID NOS: 1896
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1549
; LENGTH: 162
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-033-528-1549

Query Match
Best Local Similarity 15.1%; Score 52.4; DB 13; Length 162;
Matches 73; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

QY 4 AACATCACTGTCTTCCAGAGTACCAACACTGTCTGATGATGCGGCCCTCTTAGA 63
Db 39 AACATCACTGTCTTCCAGATTACCGACACTGTCTGACCAATACTGA-CCCTCTTTA 97
QY 64 CCTCACCACCGGGACACATGCTTCCGGTA 93
Db 98 CCTCGCCACCGGGACACACGGCTCTCTGTA 127

RESULT 12
US-10-066-543-3022/c
; Sequence 3022, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indrias, Carol Yoseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Margarita

```

```

; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.563
; CURRENT APPLICATION NUMBER: US/10/066,543
; CURRENT FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2878
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-066-543-3022

Query Match
Best Local Similarity 15.1%; Score 52.4; DB 14; Length 477;
Matches 73; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

QY 4 AACATCACTGTCTTCCAGAGTACCAACACTGTCTGATGATGCGGCCCTCTTAGA 63
Db 162 AACATCACTGTCTTCCAGATTACCGACACTGTCTGACCAATACTGA-CCCTCTTTA 104
QY 64 CCTCACCACCGGGACACATGCTTCCGGTA 93
Db 103 CCTCGCCACCGGGACACACGGCTCTCTGTA 74

RESULT 13
US-10-066-543-2878/c
; Sequence 2878, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indrias, Carol Yoseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Margarita
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.563
; CURRENT APPLICATION NUMBER: US/10/066,543
; CURRENT FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2878
; LENGTH: 549
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 526
; OTHER INFORMATION: n = A,T,C or G
US-10-066-543-2878

Query Match
Best Local Similarity 15.1%; Score 52.4; DB 14; Length 549;
Matches 73; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

QY 4 AACATCACTGTCTTCCAGAGTACCAACACTGTCTGATGATGCGGCCCTCTTAGA 63
Db 162 AACATCACTGTCTTCCAGATTACCGACACTGTCTGACCAATACTGA-CCCTCTTTA 104
QY 64 CCTCACCACCGGGACACATGCTTCCGGTA 93
Db 103 CCTCGCCACCGGGACACACGGCTCTCTGTA 74

```

RESULT 15
US-10-066-543-2864
; Sequence 2864, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yudiu
; APPLICANT: Pyie, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indirias, Carol Yoseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Margarita
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.563
; CURRENT APPLICATION NUMBER: US/10/066.543
; CURRENT FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2864
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Homo sapiens

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 23, 2003, 09:18:57 ; Search time 460.915 Seconds
(without alignments)
10917.162 Million cell updates/sec

Title: US-09-717-321A-1

Perfect score: 123

Sequence: 1 caattgaaaaagtgttgc.....ctcacagaccacaaagggtacc 123

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rtd.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	123	100.0	123	6	AX163737	AX163737 Sequence
2	123	100.0	348	6	AX163753	AX163753 Sequence
3	123	100.0	1266	6	AX163738	AX163738 Sequence
4	122	99.2	217700	2	AC106124	AC106124 Rattus no
5	107.6	87.5	2281	10	BC051053	BC051053 Mus muscu
6	107.6	87.5	2319	10	BC003828	BC003828 Mus muscu
7	107.6	87.5	269081	2	AC068493	AC068493 Mus muscu
8	106	86.2	192498	2	AC105979	AC105979 Mus muscu
9	106	86.2	230015	2	AC132602	AC132602 Mus muscu
10	100.2	81.5	238121	2	AC133022	AC133022 Rattus no
11	100.2	81.5	239768	2	AC112582	AC112582 Rattus no
12	100.2	81.5	245468	2	AC130746	AC130746 Rattus no
13	87.2	70.9	240973	2	AC123247	AC123247 Rattus no
14	87.2	70.9	241048	2	AC129824	AC129824 Rattus no
15	79.6	64.7	352	6	AX163754	AX163754 Sequence
16	79.6	64.7	455	11	G28995	G28995 human STS S
17	79.6	64.7	1232	6	AX408985	AX408985 Sequence
18	79.6	64.7	1232	9	HUMPO28T9	D25274 Homo sapien
19	79.6	64.7	2302	9	BC050687	BC050687 Homo sapi
20	79.6	64.7	2315	9	AK054993	AK054993 Homo sapi
21	79.6	64.7	28567	9	HSAL32695	HSAL32695 Homo sapi
22	79.6	64.7	212827	9	AC009412	AC009412 Homo sapi
23	77.6	63.1	262	11	G31709	G31709 SWS2233 Er
24	74.8	60.8	5544	9	AF542527	AF542527 Homo sapi
25	74.8	60.8	87616	2	AC139405	AC139405 Homo sapi
26	74.8	60.8	137625	9	AC104663	AC104663 Homo sapi
27	50	40.7	170839	2	AC133467	AC133467 Mus muscu
28	48	39.0	101584	9	CNS01D85	AL121655 BAC seque
29	48	39.0	155943	9	AC012364	AC012364 Homo sapi
30	38.2	31.1	110816	9	AC002404	AC002404 Human Chr
31	33.2	27.0	1485	5	AF029250	AF029250 Danio rer
32	33.2	27.0	1647	5	BC042319	BC042319 Danio rer
33	33.2	27.0	142171	2	EX248228	EX248228 Danio rer
34	33.2	27.0	142171	2	EX248228	EX248228 Danio rer
35	33.2	27.0	223350	2	EX465853	EX465853 Danio rer
36	33.2	27.0	223350	2	EX465853	EX465853 Danio rer
37	33	26.8	165551	2	AC106841	AC106841 Mus muscu
38	32.4	26.3	152747	9	AC092576	AC092576 Homo sapi
39	32.4	26.3	237985	2	AC125754	AC125754 Rattus no
40	32.4	26.3	247478	2	AC097964	AC097964 Rattus no
41	31.4	25.5	605	6	AX341407	AX341407 Sequence
42	31.2	25.4	255150	2	AC103167	AC103167 Rattus no
43	31	25.2	174082	2	AC116798	AC116798 Mus muscu
44	31	25.2	224777	2	AC138109	AC138109 Mus muscu
45	30.8	25.0	156879	10	AL626786	AL626786 Mouse DNA

ALIGNMENTS

RESULT 1
AX163737
LOCUS AX163737
DEFINITION Sequence 1 from Patent WO0138579.
ACCESSION AX163737
VERSION AX163737.1 GI:14544843
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1
REFERENCE
AUTHORS Gould-Rothberg, B.E., Dipippo, V.A., Ramsah, T.M. and Gerwein, R.W.
TITLE Method of identifying toxic agents using nsaid-induced differential

Pred. No. is the number of results predicted by chance to have a

```

gene expression in liver
Patent: WO 0138579-A 1 31-MAY-2001;
Curagen Corporation (US)
FEATURES
Source
1. .123
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
36 a 28 c 30 g 29 t
BASE COUNT
ORIGIN
Query Match 100.0%; Score 123; DB 6; Length 123;
Best Local Similarity 100.0%; Pred. No. 3.4e-32;
Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAATTGAAAAAGTTTGTCTTAGTGGTCGAAAGGCCCAACACTGTGTCTTGGCAGTGAG 60
Db 1 CAATTGAAAAAGTTTGTCTTAGTGGTCGAAAGGCCCAACACTGTGTCTTGGCAGTGAG 60
QY 61 TTAGGTTGTACAGAACGGCGTTAGCACTAGCGCTTGACAGAACCTCACAGACCCAAAGGT 120
Db 61 TTAGGTTGTACAGAACGGCGTTAGCACTAGCGCTTGACAGAACCTCACAGACCCAAAGGT 120
QY 121 ACC 123
Db 121 ACC 123
RESULT 2
AX163753/c
LOCUS AX163753 348 bp DNA linear PAT 22-JUN-2001
DEFINITION Sequence 17 from Patent WO0138579.
ACCESSION AX163753
VERSION AX163753.1 GI:14544859
KEYWORDS
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 Gould-Rothberg,B.E., Dipippo,V.A., Ramseh,T.M. and Gerwein,R.W.
Method of identifying toxic agents using nsaid-induced differential
gene expression in liver
Patent: WO 0138579-A 17 31-MAY-2001;
Curagen Corporation (US)
FEATURES
Source
1.348
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
103 a 76 c 67 g 102 t
BASE COUNT
ORIGIN
Query Match 100.0%; Score 123; DB 6; Length 348;
Best Local Similarity 100.0%; Pred. No. 3.7e-32;
Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY 1 CAATTGAAAAAGTTTGTCTTAGTGGTCGAAAGGCCCAACACTGTGTCTTGGCAGTGAG 60
Db 212 CAATTGAAAAAGTTTGTCTTAGTGGTCGAAAGGCCCAACACTGTGTCTTGGCAGTGAG 153
QY 61 TTAGGTTGTACAGAACGGCGTTAGCACTAGCGCTTGACAGAACCTCACAGACCCAAAGGT 120
Db 152 TTAGGTTGTACAGAACGGCGTTAGCACTAGCGCTTGACAGAACCTCACAGACCCAAAGGT 93
QY 121 ACC 123
Db 92 ACC 90
RESULT 3
AX163738

```


Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M., Hollins,B., Howells,S., Huiy,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,D., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensheva,L., Louisegh,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwakoelameh,O., Okwuonu,G., Olarpunsaagon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfankoch,C., Puzo,M., Quiroz,J., Rachin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorrelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasan,D., Waidron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhauser,A., Weis,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G., and Gibbs,R.A.

Direct Submission
Unpublished
2 (bases 1 to 217700)
Worley,K.C.

Direct Submission
Submitted (12-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 217700)

Rat Genome Sequencing Consortium.
Direct Submission
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On May 10, 2003 this sequence version replaced gi:25094619.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GHUL
Center clone name: CH230-119E11
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 185725 bases at least Q40
Consensus quality: 190140 bases at least Q30
Consensus quality: 192842 bases at least Q20
Estimated insert size: 194758; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved

* 1 109797: contig of 109797 bp in length
* 109798 109897: gap of unknown length
* 109898 166934: contig of 57037 bp in length
* 166935 167034: gap of unknown length
* 167035 176384: contig of 9350 bp in length
* 176385 176484: gap of unknown length
* 176485 206327: contig of 29843 bp in length
* 206328 206427: gap of unknown length
* 206428 207561: contig of 1034 bp in length
* 207562 207561: gap of unknown length
* 207562 208607: contig of 1046 bp in length
* 208608 208707: gap of unknown length
* 208708 209908: contig of 1201 bp in length
* 209909 210008: gap of unknown length
* 210009 211583: contig of 1575 bp in length
* 211584 211683: gap of unknown length
* 211684 213649: contig of 1966 bp in length
* 213650 213749: gap of unknown length
* 213750 215115: contig of 1366 bp in length
* 215116 215216: gap of unknown length
* 215216 217700: contig of 2485 bp in length.

FEATURES

source

BASE COUNT 51384 a 45640 c 45954 g 51476 t 23246 others
ORIGIN

Query Match 99.2%; Score 122; DB 2; Length 217700;
Best Local Similarity 100.0%; Pred. NO. 1.4e-31;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATTGAAAAGTTTGTCTAGTGTGCGAAGGCCCAACACTGTGTCTTGCAGTGAG 60
Db 64562 CAATTGAAAAGTTTGTCTAGTGTGCGAAGGCCCAACACTGTGTCTTGCAGTGAG 64503
QY 61 TTAGTTGTACAGACGGCGTTAGCAGCTTGACAGAACTCAGACACCAAGGT 120
Db 64502 TTAGTTGTACAGACGGCGTTAGCAGCTTGACAGAACTCAGACACCAAGGT 64443
QY 121 AC 122
Db 64442 AC 64441

RESULT 5
BC051053/c
LOCUS
DEFINITION Mus musculus, clone MGC:58966 IMAGE:5038182, mRNA linear ROD 14-APR-2003
ACCESSION BC051053
VERSION BC051053.1 GI:29835221
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2281)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (11-APR-2003) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc_mgc@nigr.nih.gov
Akhter, N., Ayle, K., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Blakesley, R.W., Bouffard, G.G., Breen, K., Gupta, J., Haghighi, P., Dietrich, N.L., Granite, S., Guan, X., Gupta, P., Legaspi, R., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Maduro, Q.L., Masliello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

REMARK
COMMENT

BC003828/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

BC003828
MGC:6235 IMAGE:3593957, complete cds.
BC003828
MGC.
BC003828.1 GI:13277917
MGC.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2319)
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stachek, M., Soares, M.B., Bonaldo, M.F., Usdin, T.B., Toshiyuki, S., Scheetz, T.E., Brownstein, M.J., Ustin, T.B., Loquellano, N.A., Peters, G.J., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaby, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettner, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickinson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Small, D.E., Scherch, A., Schein, J.E., Jones, S.J. and Marra, M.A. M.A. full-length generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
12477932
2 (bases 1 to 2319)
Direct Submission
Submitted (28-FEB-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www.shgc.stanford.edu>
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAX Plate: 11 Row: e Column: 22.
Location/Qualifiers
1. .2319
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="MGC:6235 IMAGE:3593957"
/tissue_type="Mammary tumor. Metallothionien-TGF alpha model. 10 month old virgin mouse. Taken by biopsy."
/clone_lib="NCI CGAP Mam1"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
1. .2319
/gene="Racl"
/note="synonym: D5Erttd559e"

FEATURES
source
gene

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAX Plate: 108 Row: f Column: 9
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, GenomeScan gene prediction, Similarity but not identity to protein.
Location/Qualifiers
1. .2281
/organism="Mus musculus"
/mol_type="mRNA"
/strain="CZECH II"
/db_xref="taxon:10090"
/clone="MGC:58966 IMAGE:5038182"
/tissue_type="Mammary tumor metastasized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMTV."
/clone_lib="NCI CGAP_Lu29"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
197. .775
/codon_start=1
/product="Unknown (protein for MGC:58966)"
/protein_id="AAH51053.1"
/db_xref="GI:29835222"
/translation="MQAKVGVVGDGAVGKTCLLISVYTNAPFGEVLPVFDNYSANV MVDKPNVLGLMDTAGEDYDLRLPLSPYQTDVFLICFSLVSPASFNPAKWPVEVR HHCNPPIILVGTGLDLRDDKDTIEKLKKLTPITYPQGLAWAKEIGAVKYLECSAL TORGLKTVFDEARAVLCPPPPVKKRKCKILL"
BASE COUNT 568 a 569 c 504 g 640 t
ORIGIN

Query Match 87.5%; Score 107.6; DB 10; Length 2281;
Best Local Similarity 92.6%; Pred. No. 1.1e-26;
Matches 113; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 1 CAATTGAAGAAAGTTGTTCTAGTGGTCGAAGGCCCAACACTGCTTCTTGCCAGTGAG 60
DB 2170 CAATTGAAGAAAGTTGTTCTAGTGGTTGAAGGCCCAACACTGCTTCTTGCCAGTGAG 2111
QY 61 TTAGTTGTGACAGAACCGCGTTAGCACTAGCGCTTGACAGAACCTCAGACCCCAAGGT 120
DB 2110 TTAGTTGTGACAGACACGCGTTAGCACTAGCAGTTGACAGAACCTCAGACCCCAAGGA 2051
QY 121 AC 122
DB 2050 AC 2049
RESULT 6

```
/db_xref="LocusID:19353"
/db_xref="MGI:97845"
198 - 776
/codon_start=1
/product="rac1 protein"
/protein_id="AAH03828.1"
/db_xref="GI:13277918"
/db_xref="LocusID:19353"
/translation="MQAIKVVGDGAVGKTCILLISYTNAPGGEYPTVPDYNVSNV
MVQGKPNVLGMDTAGEDDRRLPLSYPTDPLFCISLVSPASFENVRKAYPEVR
HHCPTDILVGTGLDRDDKDTLEKLEKELKLTPTIYPOGLAMAKAIGCAVKYLECSAL
TQRLKTVDFEALRVLCPFPVKKRKKCLLL"
BASE COUNT 603 a 573 c 503 g 640 t
ORIGIN
Query Match 87.5%; Score 107.6; DB 10; Length 2319;
Best Local Similarity 92.6%; Pred. No. 1.1e-26;
Matches 113; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 1 CAATTGAAAAAGTTTCTTCTAGTGGTCGAAAGGCCCAACACTGTGTTCTTGCAGTGAG 60
Db 2173 CAATTGAAAAAGTTTCTTCTAGTGGTTGAAAGGCCCAACACTGTGTTCTTGCAGTGAG 2114
QY 61 TTAGTGTGTACAAACGGCGTTAGCACTAGCGTTGACAGAACTCAGACCCCAAGGT 120
Db 2113 TTAGTGTGTACAGCAGCGTTAGCACTAGCGTTGACAGAACTCAGACCCCAAGGA 2054
QY 121 AC 122
Db 2053 AC 2052
RESULT 7
AC068493/c
LOCUS 269081 bp DNA linear HTG 15-MAY-2002
DEFINITION Mus musculus clone RP23-76K1 strain C57BL6/J, WORKING DRAFT
ACCESSION AC068493
VERSION 26
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE Mus musculus
ORGANISM Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 269081)
AUTHORS Montgomery,K.T., Grills,G., Han,J., Lee,E., Long,J., Pomerantz,R.,
Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,
Gordon,M., Goltz,J.S. and Kucherlapati,R.
High Throughput Mouse Sequencing
Unpublished
2 (bases 1 to 269081)
AUTHORS Montgomery,K.T., Grills,G., Han,J., Lee,E., Long,J., Pomerantz,R.,
Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,
Gordon,M., Goltz,J.S. and Kucherlapati,R.
Direct Submission
TITLE Submitted (03-MAY-2000) Department of Molecular Genetics, Albert
JOURNAL Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
Brux, NY 10461, USA
COMMENT On Aug 11, 2001 this sequence version replaced gi:14993654.
-----Genome Center
Center: Harvard Partners Genome Center
Center Code: HPGC
Web site: http://www.hpcgg.org/Sequence/mouse.html
Contact: hpcgcmendel.mgh.harvard.edu
-----Summary Statistics
Center project name: ABN
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 100%
*Consensus quality: 256302 at least Q20
*Consensus quality: 254124 at least Q30
*Consensus quality: 250276 at least Q40
Estimated insert size: agarose-PF - N/A
**Estimated insert size: 268581 - sum-of-contigs
```

Quality coverage: agarose-PF - N/A
Quality coverage: 6.2 x in Q20 bases; sum-of-contigs estimation

* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 51730: contig of 51730 bp in length
* 51731: gap of unknown length
* 51751: contig of 30673 bp in length
* 82423: gap of unknown length
* 82424: gap of unknown length
* 118158: contig of 35715 bp in length
* 118159: gap of unknown length
* 118178: gap of unknown length
* 118179: contig of 21263 bp in length
* 139441: contig of 21263 bp in length
* 139442: gap of unknown length
* 139461: gap of unknown length
* 160638: contig of 21177 bp in length
* 160639: gap of unknown length
* 160659: contig of 16953 bp in length
* 176611: contig of 16953 bp in length
* 176612: gap of unknown length
* 177631: contig of 13360 bp in length
* 177632: gap of unknown length
* 190992: contig of 9117 bp in length
* 191012: contig of 9117 bp in length
* 200129: gap of unknown length
* 200148: gap of unknown length
* 211882: contig of 11734 bp in length
* 211902: gap of unknown length
* 220773: contig of 8871 bp in length
* 220774: gap of unknown length
* 229643: contig of 8850 bp in length
* 229644: gap of unknown length
* 229663: gap of unknown length
* 238339: contig of 8676 bp in length
* 238340: gap of unknown length
* 238359: gap of unknown length
* 244868: contig of 6489 bp in length
* 244869: gap of unknown length
* 249725: contig of 4857 bp in length
* 249745: gap of unknown length
* 252523: contig of 2778 bp in length
* 252524: gap of unknown length
* 252543: gap of unknown length
* 252544: contig of 4977 bp in length
* 257520: gap of unknown length
* 257521: contig of 2856 bp in length
* 257541: gap of unknown length
* 260397: contig of 2293 bp in length
* 260417: gap of unknown length
* 262709: contig of 2293 bp in length
* 262710: gap of unknown length
* 262730: contig of 1205 bp in length
* 263935: gap of unknown length
* 263955: contig of 242 bp in length
* 264197: gap of unknown length
* 264217: contig of 413 bp in length
* 264629: gap of unknown length
* 264630: contig of 959 bp in length
* 265608: gap of unknown length
* 265609: contig of 1307 bp in length
* 266935: gap of unknown length
* 266936: contig of 1307 bp in length
* 266955: gap of unknown length
* 266956: contig of 1380 bp in length
* 268335: gap of unknown length
* 268355: contig of 194 bp in length
* 268356: gap of unknown length
* 268550: contig of 512 bp in length.
* 268551: contig of 512 bp in length.

FEATURES
Source

1. 269081
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL6/J"
/db_xref="taxon:10090"
/clone="RP23-76K1"
/sex="male"
1. 51730
/note="assembly_name:Contig138"
51751. 82423
misc_feature
misc_feature

```

/misc_feature /note="assembly_name:Contig137" 82444..118158
/misc_feature /note="assembly_name:Contig136" 118179..139441
/misc_feature /note="assembly_name:Contig135" 139462..160638
/misc_feature /note="assembly_name:Contig134" 160859..177611
/misc_feature /note="assembly_name:Contig133" 177632..190991
/misc_feature /note="assembly_name:Contig132" 191012..200128
/misc_feature /note="assembly_name:Contig131" 200149..211862
/misc_feature /note="assembly_name:Contig130" 211903..220773
/misc_feature /note="assembly_name:Contig129" 220794..229643
/misc_feature /note="assembly_name:Contig128" 229664..238339
/misc_feature /note="assembly_name:Contig127" 238360..244848
/misc_feature /note="assembly_name:Contig126" 244869..249725
/misc_feature /note="assembly_name:Contig125" 249746..252523
/misc_feature /note="assembly_name:Contig124" clone end:SP6
vector side:left
/misc_feature 252544..257520
/misc_feature /note="assembly_name:Contig123" 257541..260396
/misc_feature /note="assembly_name:Contig122" 260417..262709
/misc_feature /note="assembly_name:Contig121" 262730..263934
/misc_feature /note="assembly_name:Contig120" 263955..264196
/misc_feature /note="assembly_name:Contig119" 264217..264629
/misc_feature /note="assembly_name:Contig118" 264650..265608
/misc_feature /note="assembly_name:Contig117" 265629..266935
/misc_feature /note="assembly_name:Contig116" 266956..268335
/misc_feature /note="assembly_name:Contig115" 268356..268549
/misc_feature /note="assembly_name:Contig114" 268570..269081
/misc_feature /note="assembly_name:Contig113" 269102..269614
BASE COUNT 69726 a 66027 c 64672 g 67724 t 932 others
ORIGIN
Query Match 87.5% Score 107.6; DB 2; Length 269081;
Best Local Similarity 92.6%; Pred No. 1.6e-26;
Matches 113; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 1 CAAATTGAAAAGTTGTTCTAGTGGTCGAAGGCCCAACACTGTGTTCTTGCAGTGAG 60
Db 63360 CAAATTGAAAAGTTGTTCTAGTGGTCGAAGGCCCAACACTGTGTTCTTGCAGTGAG 63301
QY 61 TTAGGTTGTACAGACGGGTTAGCCTAGCGCTTGACAGAACTCAGACACCAAGGT 120
Db 63300 TTAGGTTGTACAGACGGGTTAGCCTAGCGCTTGACAGAACTCAGACACCAAGGT 120
QY 121 AC 122
Db 63240 AC 63239

```

RESULT 8
AC105979

```

LOCUS AC105979 192498 bp DNA linear HTG 21-OCT-2002
DEFINITION Mus musculus clone RP24-301E12, WORKING DRAFT SEQUENCE, 9 ordered
pieces.
ACCESSION AC105979
VERSION AC105979.4 GI:24182089
KEYWORDS HTG; HTGS PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Birren,B., Nusbaum,C. and Lander,E.
1 (bases 1 to 192498)
Mus musculus, clone RP24-301E12
Unpublished
2 (bases 1 to 192498)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepe,J., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrum,J.,
Meneus,L., Mihova,T., Mlenka,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (11-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 192498)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepe,J., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
Matthews,C., McCarthy,M., Meldrum,J., Meneus,L., Mihova,T.,
Mlenka,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupack,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Talamas,J., Testaye,S., Theodore,J., Topham,K.,
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (21-OCT-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Oct 21, 2002 this sequence version replaced gi:20149424.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L20199

```

```

LOCUS AC105979 192498 bp DNA linear HTG 21-OCT-2002
DEFINITION Mus musculus clone RP24-301E12, WORKING DRAFT SEQUENCE, 9 ordered
pieces.
ACCESSION AC105979
VERSION AC105979.4 GI:24182089
KEYWORDS HTG; HTGS PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Birren,B., Nusbaum,C. and Lander,E.
1 (bases 1 to 192498)
Mus musculus, clone RP24-301E12
Unpublished
2 (bases 1 to 192498)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepe,J., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrum,J.,
Meneus,L., Mihova,T., Mlenka,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (11-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 192498)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepe,J., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
Matthews,C., McCarthy,M., Meldrum,J., Meneus,L., Mihova,T.,
Mlenka,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupack,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Talamas,J., Testaye,S., Theodore,J., Topham,K.,
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (21-OCT-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Oct 21, 2002 this sequence version replaced gi:20149424.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L20199

```

```

Center clone name: 301_E_12
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 190070 bases at least Q40
Consensus quality: 191243 bases at least Q30
Consensus quality: 191514 bases at least Q20
Insert size: 163000; agarose-fp
Insert size: 191698; sum-of-contigs
Quality coverage: 11.6 in Q20 bases; agarose-fp
Quality coverage: 9.9 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 1061: contig of 1061 bp in length
* 1062 1161: gap of 100 bp
* 1182 5714: contig of 4553 bp in length
* 5715 5814: gap of 100 bp
* 5815 11422: contig of 5608 bp in length
* 11423 11522: gap of 100 bp
* 11523 17600: contig of 6078 bp in length
* 17601 17700: gap of 100 bp
* 17701 30157: contig of 12457 bp in length
* 30158 30257: gap of 100 bp
* 30258 38787: contig of 8530 bp in length
* 38788 38887: gap of 100 bp
* 38888 99444: contig of 60557 bp in length
* 99445 99544: gap of 100 bp
* 99545 143263: contig of 43719 bp in length
* 143264 143363: gap of 100 bp
* 143364 192498: contig of 49135 bp in length.
*
* Location/Qualifiers
* 1..192498
* /organism="Mus musculus"
* /mol_type="genomic DNA"
* /db_xref="taxon:10090"
* /clone_lib="RP24-301E12"
* /clone_lib="RPCI-24 Male Mouse BAC"
*
* misc_feature
* 1..1061
* /note="assembly_fragment"
* 1162..5714
* /note="assembly_fragment"
* 5815..11422
* /note="assembly_fragment"
* 11523..17600
* /note="assembly_fragment"
* 17701..30157
* /note="assembly_fragment"
* 30258..38787
* /note="assembly_fragment"
* 38888..99444
* /note="assembly_fragment"
* 99545..143263
* /note="assembly_fragment"
* 143364..192498
* /note="assembly_fragment"
* clone end:77
* vector side:right"
*
BASE COUNT 53596 a 42471 c 43252 g 52379 t 800 others
ORIGIN
Query Match 86.2%; Score 106; DB 2; Length 192498;
Best Local Similarity 91.8%; Pred. No. 5,7e-26;
Matches 112; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```

```

Qy 1 CAATTGAAAAAGTTTCTTCTAGTCGTCGAAAGCCCAACACTGTGTCTTCCAGTGAG 60
Db 90351 CACTGAAAAAGTTTCTTCTAGTCGTCGAAAGCCCAACACTGTGTCTTCCAGTGAG 90410
Qy 61 TTAGTTGTACAGAACGCGCTTAGCCTAGCGCTTGACAGAACCTCAGACCCCAAGGT 120
Db 90411 TTAGTTGTACAGACGACGACATTAGCCTAGCAGTTGACAGAACCTCAGACCCCAAGGA 90470
Qy 121 AC 122
Db 90471 AC 90472

RESULT 9
AC132602 230015 bp DNA linear HTG 01-SEP-2002
LOCUS Mus musculus chromosome UNK clone RP23-477C13, WORKING DRAFT
DEFINITION SEQUENCE, 10 unordered pieces.
ACCESSION AC132602
VERSION AC132602.1 GI:22597688
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 230015)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 230015)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: M.BA0477C13
----- Summary Statistics -----
Sequencing vector: M13; 0%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 225083 bases at least Q40
Consensus quality: 225730 bases at least Q30
Consensus quality: 226267 bases at least Q20
Insert size: 202000; agarose-fp
Insert size: 228405; sum-of-contigs
Quality coverage: 13.11 in Q20 bases; agarose-fp
Quality coverage: 9.64 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 710: contig of 710 bp in length
* 711 810: gap of unknown length
* 811 18316: contig of 17506 bp in length
* 18317 18416: gap of unknown length
* 18417 57380: contig of 38964 bp in length
* 57381 57480: gap of unknown length
* 57481 118173: contig of 60693 bp in length
* 118174 118273: gap of unknown length

```

```

* 118274 201518: contig of 83245 bp in length
* 201519 201618: gap of unknown length
* 201619 202801: contig of 1183 bp in length
* 202802 202901: gap of unknown length
* 202902 204714: contig of 1813 bp in length
* 204715 204814: gap of unknown length
* 204815 208091: contig of 3277 bp in length
* 208092 208191: gap of unknown length
* 208192 213082: contig of 4891 bp in length
* 213083 213182: contig of unknown length
* 213183 230015: contig of 16833 bp in length.

```

FEATURES

```

source
1. 230015
  /organism="Mus musculus"
  /mol_type="genomic DNA"
  /db_xref="taxon:10090"
  /chromosome="X"
  /clone="RP23-477C13"
  /notes="assembly_name:Contig1"
misc_feature
811..18316
  /note="assembly_name:Contig10"
misc_feature
18417..57380
  /note="assembly_name:Contig11"
misc_feature
57481..118173
  /note="assembly_name:Contig12"
misc_feature
118274..201518
  /note="assembly_name:Contig13"
misc_feature
201619..202801
  /note="assembly_name:Contig13"
misc_feature
202902..204714
  /note="assembly_name:Contig13"
misc_feature
204815..208091
  /note="assembly_name:Contig13"
misc_feature
208192..213082
  /note="assembly_name:Contig13"
misc_feature
213183..230015
  /note="assembly_name:Contig13"
misc_feature
63025..50551 c 50278 g 65213 t 948 others
ORIGIN

```

```

Query Match      86.2%  Score 106;  DB 2;  Length 230015;
Best Local Similarity 91.8%  Pred. No. 5.8e-26;
Matches 112;  Conservative 0;  Mismatches 10;  Indels 0;  Gaps 0;

```

```

QY 1 CAATTGAAAGAGTTGTTCTAGTGGTCCGAAGGCCCAACACTGTGTTCTTGCAGTGCAG 60
   |||||
Db 109078 CAATTGAAAGAGTTGTTCTAGTGGTCCGAAGGCCCAACACTGTGTTCTTGCAGTGCAG 109137
   |||||

QY 61 TTAGTTGTACAGACGGCGTTAGCACTAGCGTTGACAGAACTCAGACACCCCAAGGT 120
   |||||
Db 109138 TTAGTTGTACAGACGGCGTTAGCACTAGCGTTGACAGAACTCAGACACCCCAAGGT 109197
   |||||

QY 121 AC 122
   ||
Db 109198 AC 109199

```

```

RESULT 10
LOCUS AC133022/c
DEFINITION Rattus norvegicus clone CH230-50516, *** SEQUENCING IN PROGRESS
***, 3 unordered pieces.
ACCESSION AC133022
VERSION 3
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 228121)
AUTHORS Muzny,D,Marie., Metzker,M, Lee., Abramson,S., Adams,C., Alder,J.,

```

Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.I., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Georegeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.B., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Loulseg, H., Lozada, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmood, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, B., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Polindexter, A., Popovic, D., Primus, E., Pu, L., Pu, L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scheter, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sittler, C.D., Smajis, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Wortley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 228121)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (06-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 228121)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 20, 2002 this sequence version replaced gi:23908439.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

On Sep 23, 2002 this sequence version replaced gi:21738449.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GRAL

Center clone name: CH230-257N6

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 196112 bases at least Q40

Consensus quality: 198658 bases at least Q30

Consensus quality: 200655 bases at least Q20

Estimated insert size: 216180; sum-of-contigs estimation

Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 12 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 120874: contig of 120874 bp in length

* 120875 120974: gap of unknown length

* 120975 173485: contig of 52511 bp in length

* 173486 173585: gap of unknown length

* 173586 192533: contig of 18948 bp in length

* 192534 192633: gap of unknown length

* 192634 193678: contig of 1045 bp in length

* 193679 193778: gap of unknown length

* 193779 194803: contig of 1025 bp in length

* 194804 194903: gap of unknown length

* 194904 197578: contig of 2675 bp in length

* 197579 197678: gap of unknown length

* 197679 200980: contig of 3302 bp in length

* 200981 201080: gap of unknown length

* 201081 214297: contig of 13217 bp in length

* 214298 214397: gap of unknown length

* 214398 217290: contig of 2893 bp in length

* 217291 217390: gap of unknown length

* 217391 226678: contig of 9288 bp in length

* 226679 226778: gap of unknown length

* 226779 236078: contig of 9299 bp in length

* 236078 236177: gap of unknown length

* 236178 239768: contig of 3591 bp in length.

* Location/Qualifiers

1. 239768

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-257N6"

BASE COUNT 62543 a 42292 c 41173 g 55976 t 37784 others

ORIGIN

Query Match 81.58; Score 100.2; DB 2; Length 239768;

Best Local Similarity 89.3%; Pred. No. 6.4e-24;

Matches 108; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 2 ATTGAAAAAGTTTCTTAGTGGTCGAAAGCCCAACACTGTGTTCTTGCAGTAGT 61

Db 19799 AAGAAAAAAGTTTCTTAGTGGTTGAAGGCCCAACACTATATCTTGCAGTAGT 18858

QY 62 TAGGTTGTACAGAACGGCGTTAGCACTAGCGTTGACAGAACCTCAGACCCAAAGGTA 121

Db 18859 TTGTTTACAGAAAGCGTTAGCACTAGCGTTGACAGAACCTCAGACCCAAAGGAA 18918

QY 122 C 122

Db 18919 C 18919

RESULT 12

AC130746 245468 bp DNA linear HTG 10-MAY-2003

LOCUS Rattus norvegicus clone CH230-39M21, *** SEQUENCING IN PROGRESS

DEFINITION *** 3 unordered pieces.

AC130746

AC130746.3 GI:30522353

VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.

KEYWORDS Rattus norvegicus (Norway rat)

SOURCE Rattus norvegicus

ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 245468)

AUTHORS Muzny,D.,Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Balgwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Bismwal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpthy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuewa,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K., Nwakoelmeht,O., Okwuonu,G., Olarnpungsoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L., Pu,L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Sma's,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,

Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Zhao,S., Dunn,D., von
Niederhausern,A., Weis,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.

Direct Submission

Unpublished

2 (bases 1 to 245468)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (14-AUG-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 245468)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

On May 10, 2003 this sequence version replaced gi:23814855.

The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GGTN

Center clone name: CH230-39M21

----- Summary Statistics

Assembly program: Atlas 3.0;

Consensus quality: 236558 bases at least Q40

Consensus quality: 238736 bases at least Q30

Consensus quality: 240308 bases at least Q20

Estimated insert size: 249127; sum-of-contigs estimation

Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 242398: contig of 242398 bp in length

* 242399 242498: gap of unknown length

* 242499 243549: contig of 1051 bp in length

* 243550 243649: gap of unknown length

* 243650 245468: contig of 1819 bp in length.

Location/Qualifiers

1. .245468

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-39M21"

1. .1572

/notes="wgs end extension

clone end:Sp6"

2074. .3190

misc_feature

misc_feature

/note="wgs end extension
clone end:Sp6"

4356. .5049

/note="clone boundary

clone end:Sp6

site:ECORI

end sequence:BH286156"

238293. .239134

/note="clone boundary

clone end:T7

site:ECORI

end sequence:BH286155"

BASE COUNT 70063 a 51006 c 48284 g 71652 t 4463 others

ORIGIN

Query Match 81.5%; Score 100.2; DB 2; Length 245468;

Best Local Similarity 89.3%; Pred. No. 6.4e-24;

Matches 108; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 2 AATTGAAAAAGTTTCTAGTGGTCTGAAAGCCCAACACTGTGTTCTTGCACGTGAGT 61

Db 232776 AAGAAAAAAGTTTCTAGTGGTCTGAAAGCCCAACACTATATATCTTGCACGTGAGT 232835

Qy 62 TAGTTCTGACAGAACGCGTTAGCACTAGCGCTTGACAGAACCTCAGACCCCAAGGTA 121

Db 232836 TTGTTACACAGATGCGTTAGCACTAGCGTTTGACAGAACCTCAGACCCCAAGGTA 232895

Qy 122 C 122

Db 232896 C 232896

RESULT 13

AC123247

LOCUS

AC123247 Rattus norvegicus clone CH230-151E3, WORKING DRAFT SEQUENCE, 3
unordered pieces.

AC123247

AC123247.4 GI:25138228

HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE 1 (bases 1 to 240973)

AUTHORS

Muzny,D.,Marle., Metzker,M.,Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Ayalebechi,V.,Ayagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaika,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Diyya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Evans,K.,
Rgan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Gartner,T., Garza,M.,
Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Loulsegod,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,

Mawhiney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munday, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackelme, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, P., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, E., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE JOURNAL REFERENCE AUTHORS JOURNAL

Unpublished
2 (bases 1 to 240973)
Worley, K.C.
Direct Submission
Submitted (29-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 240973)

REFERENCE AUTHORS JOURNAL

Rat Genome Sequencing Consortium.
Direct Submission
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 20, 2002 this sequence version replaced gi:23097393.

COMMENT

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GTOE
Center Clone name: CH230-151B3
----- Summary Statistics

Assembly program: Phrap; version 0.990329
Consensus quality: 230127 bases at least Q40
Consensus quality: 232322 bases at least Q30
Consensus quality: 233892 bases at least Q20
Estimated insert size: 236773; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence

* as soon as it is available and the accession number will
* be preserved.

* 1 159537: contig of 159537 bp in length
* 159538 159637: gap of unknown length
* 159638 239630: contig of 79993 bp in length
* 239631 239730: gap of unknown length
* 239731 240973: contig of 1243 bp in length.

FEATURES source

1. 240973
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-151B3"

misc_feature
1..1047

/note="wgs contig"

misc_feature
3510..5688

/note="wgs contig"

BASE COUNT 57688 a 61245 c 61509 g 54580 t 5951 others
ORIGIN

Query Match 70.9%; Score 87.2; DB 2; Length 240973;
Best Local Similarity 87.7%; Pred. No. 2.4e-19;
Matches 107; Conservative 0; Mismatches 13; Indels 2; Gaps 1;

QY 1 CAATTGAAAAAGTTTGTCTAGTGTGCAAGGCGCCCAACACTGTGTCTTGTCCAGTGAG 60
Db 203927 CAATTGAAAAAGTTTGTCTAGGCGCTTAAAGGCCCACTA--TTCTTGCAGTGAG 203984

QY 61 TTAGGTTGTACAGACGGGCTTAGCTAGCGCTTGACAGAACCTCAGACGCCCAAGGT 120
Db 203985 ATAGGTTATATAGATGGGCTCAGCACTAGCTGACAGAACCTCACAACCCAAAGGA 204044

QY 121 AC 122

Db 204045 AC 204046

RESULT 14

AC129824/c

LOCUS Rattus norvegicus clone CH230-88022, *** SEQUENCING IN PROGRESS
DEFINITION

AC129824.3 GI:23265731

VERSION HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.

KEYWORDS Rattus norvegicus (Norway rat)

SOURCE Rattus norvegicus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 241048)

REFERENCE

AUTHORS

Muzny, D., Marie, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Altschuler, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, I., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Genatragoe, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guavara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, J., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensheva, L., Lounseghed, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackeleme, O., Okwuonu, G., Olampunsaagoo, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, M., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Steimle, M., Strong, R., Sutton, A., Sveltek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G., and Gibbs, R.A.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Direct Submission

Unpublished

2 (bases 1 to 241048)

Worley, K.C.

Direct Submission

Submitted (03-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 241048)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (21-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Sep 21, 2002 this sequence version replaced gi:22538942. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GSWG

Center clone name: CH230-8022

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 215098 bases at least Q40

Consensus quality: 217421 bases at least Q30

Consensus quality: 218930 bases at least Q20

Estimated insert size: 230359; sum-of-contigs estimation

Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

***** NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.

* This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 241048: contig of 241048 bp in length.

FEATURES

source

1. 241048
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-80022"

misc_feature

1. 1324
 /note="wgs_end_extension
 clone_end:Sp6"

misc_feature

4678..6863
 /note="wgs_end_extension
 clone_end:Sp6"

misc_feature

complement(7042..7750)
 /note="clone_boundary
 clone_end:Sp6"

misc_feature

end sequence: BH317537
 181147..182998
 /note="wgs_contig"

misc_feature

234297..235118
 /note="clone_boundary
 clone_end:T7"

misc_feature

end sequence: BH317534
 238585..241048
 /note="wgs_end_extension
 clone_end:T7"

misc_feature

BASE COUNT 51891 a 56832 c 56433 g 55074 t 20819 others
 ORIGIN

Query Match

Best Local Similarity

Matches

107; Conservative

0; Mismatches

13; Indels

2; Gaps

1;

Qy

1

CAATTGAAAAAGTTTCTTCTAGTGTGCGAAAGGCCCAACACTGTCTTCTGCGAGTCTAG

60

Db

156075

CAATTGAAAAAGTTTCTTCTAGTGTGCGAAAGGCCCAACACTGTCTTCTGCGAGTCTAG

156018

Qy

61

TTAGTTGTACAGAACGGGTAGCCTAGCGTTGACAGAACCTCAGACCCCAAGGT

120

Db

156017

ATAGTTTATAGAAATGGCGTCTAGCCTAGCAGTTGACAGAACCTCAGACCCCAAGGT

155958

Qy

121

AC 122

Db

155957

AC 155956

RESULT 15

AX163754/c

LOCUS

AX163754

DEFINITION

Sequence 18 from Patent WO0138579.

ACCESSION

AX163754

VERSION

AX163754.1

GI:14544860

KEYWORDS

Homo sapiens (human)

SOURCE

Homo sapiens

ORGANISM

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1

AUTHORS

Gould-Rothberg, B.E., Dipippo, V.A., Rameh, T.M. and Gerwein, R.W.

TITLE

Method of identifying toxic agents using nsaid-induced differential

JOURNAL

Patent: WO 0138579-A 18 31-MAY-2001;

Curagen Corporation (US)

FEATURES

Location/Qualifiers

source

1. .352

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

BASE COUNT	86 a	77 c	72 g	117 t
ORIGIN				

Query Match 64.7%; Score 79.6; DB 6; Length 352;
Best Local Similarity 83.1%; Pred. No. 6.8e-17;
Matches 103; Conservative 0; Mismatches 19; Indels 2; Gaps 1;

Qy 1 CAATTGAAAAAGTTTGTCTTAGTGGTCGAAAGGCCCAACACTGTGTCTCTTGCCAGTGAG 60

D_b 216 CAATTAAATAATTCTGTTGTAGTGGCTGAAGGGTCCCACGCTGTATTCTCGCCAGTGAG 157

QY 61 TTAGGTTGTACAGAACGGCGTTAGCACTAGCGC--TTGACAGAACCTCACAGACCCAAAG 118

D_b 156 TTAAGTTGTACAGAACATCGTCAGCCTAGCACAGTTTACAGAACTTACAGACCCAAAG 97

QY 119 GTAC 122

Db 96 GAAC 93

Search completed: November 23, 2003, 14:01:45
Job time : 468.915 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 23, 2003, 10:50:08 ; Search time 9.37908 Seconds
(without alignments)
5788.425 Million cell updates/sec

Title: US-09-717-321A-1

Perfect score: 123

Sequence: 1 caattgaaaaagttgttc.....ctcacagacccaagggtacc 123

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA:*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
- 5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	31.4	25.5	605	3	US-09-385-982-483	Sequence 483, App
2	28.8	23.4	202001	4	US-09-734-674-3	Sequence 3, Appli
3	26.8	21.8	6827	4	US-08-961-527-60	Sequence 60, Appl
4	26.6	21.6	81001	4	US-09-750-580-1	Sequence 1, Appli
5	26.4	21.5	786	4	US-09-252-991A-778	Sequence 778, App
6	26.4	21.5	900	4	US-09-252-991A-638	Sequence 638, App
7	26.4	21.5	1035	4	US-09-457-066-6	Sequence 6, Appli
8	26.4	21.5	1035	4	US-09-706-968-6	Sequence 6, Appli
9	26.4	21.5	1572	4	US-09-328-352-2269	Sequence 2269, Ap
10	26.4	21.5	536165	4	US-09-214-808-1	Sequence 1, Appli
11	25.6	20.8	3033	2	US-08-462-880B-1	Sequence 1, Appli
12	25.6	20.8	3033	3	US-08-462-090-1	Sequence 1, Appli
13	25.6	20.8	3033	3	US-08-463-461-1	Sequence 1, Appli
14	25.4	20.7	1664976	4	US-08-916-421B-1	Sequence 1, Appli
15	25.2	20.5	49272	1	US-08-614-770A-1	Sequence 1, Appli
16	25	20.3	550	3	US-08-705-771-1	Sequence 1, Appli
17	25	20.3	786	5	PCT-US98-08295-1	Sequence 1, Appli
18	24.8	20.2	1572	4	US-09-328-352-2268	Sequence 2268, Ap
19	24.6	20.0	3095	1	US-08-452-083-1	Sequence 1, Appli
20	24.6	20.0	80595	3	US-09-078-294-3	Sequence 3, Appli
21	24.6	20.0	4403765	3	US-09-103-840A-2	Sequence 2, Appli
22	24.6	20.0	4411529	3	US-09-103-840A-1	Sequence 1, Appli
23	24.4	19.8	1335	4	US-09-328-352-2337	Sequence 2337, Ap
24	24.2	19.7	834	4	US-09-252-991A-3547	Sequence 3547, Ap
25	24.2	19.7	1176	4	US-09-252-991A-3540	Sequence 3540, Ap
26	24.2	19.7	2832	4	US-09-252-991A-3523	Sequence 3523, Ap
27	24.2	19.7	9306	4	US-09-453-702B-231	Sequence 231, App

C 28	24.2	19.7	36470	4	US-08-311-731A-123	Sequence 123, App
C 29	24	19.5	231	4	US-09-328-352-380	Sequence 380, App
C 30	24	19.5	275	4	US-09-364-206-33	Sequence 33, Appl
C 31	24	19.5	475	4	US-09-312-283C-367	Sequence 367, App
C 32	24	19.5	861	4	US-09-252-991A-665	Sequence 665, App
C 33	24	19.5	1971	4	US-09-107-532A-1429	Sequence 1429, Ap
C 34	23.8	19.3	1110	4	US-09-564-535D-6	Sequence 6, Appli
C 35	23.8	19.3	2326	4	US-09-634-238-28	Sequence 28, Appli
C 36	23.6	19.2	1137	4	US-09-252-991A-5529	Sequence 5529, Ap
C 37	23.6	19.2	1545	4	US-09-252-991A-5553	Sequence 5553, Ap
C 38	23.6	19.2	1914	4	US-09-252-991A-5670	Sequence 5670, Ap
C 39	23.6	19.2	112132	4	US-08-741-150-3	Sequence 3, Appli
C 40	23.4	19.0	228	1	US-08-250-958-3	Sequence 3, Appli
C 41	23.4	19.0	231	5	PCT-US95-00605-2	Sequence 2, Appli
C 42	23.4	19.0	588	3	US-09-385-982-133	Sequence 133, App
C 43	23.4	19.0	725	4	US-09-591-992-3	Sequence 3, Appli
C 44	23.4	19.0	741	3	US-07-927-391-25	Sequence 25, Appl
C 45	23.2	18.9	376	4	US-09-702-705-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1

US-09-385-982-483
; Sequence 483, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: PRODUCTS: II
; FILE REFERENCE: CCNDA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 483
; LENGTH: 605
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(605)
; OTHER INFORMATION: n = A,T,C or G

US-09-385-982-483

Query Match 25.5%; Score 31.4; DB 3; Length 605;
Best Local Similarity 85.5%; Pred. No. 0.026; Indels 2; Gaps 1;
Matches 47; Conservative 0; Mismatches 6;

Oy 70 ACAGACGGCTTAGCCTAGCGC--TTGACAGAACCTCAGACCCAAAGGTAC 122
|||
Db 1 ACAGACATCGTCAGCAGCTAGCAGTTTACAGAACCTCAGACCCAAAGGAC 55

RESULT 2

US-09-734-674-3/c
; Sequence 3, Application US/09734674
; Patent No. 6498022

; GENERAL INFORMATION:

; APPLICANT: WEI, Ming-Hui et al

; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF

; FILE REFERENCE: CL001018

; CURRENT APPLICATION NUMBER: US/09/734,674

; CURRENT FILING DATE: 2000-12-13


```
OTHER INFORMATION: 20-841-149 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 45442
OTHER INFORMATION: 20-842-115 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 77058
OTHER INFORMATION: 20-853-415 : polymorphic base C or T
NAME/KEY: primer:bind
LOCATION: 929..949
OTHER INFORMATION: 20-828.pu
NAME/KEY: primer:bind
LOCATION: 1357..1377
OTHER INFORMATION: 20-828.rp complement
NAME/KEY: primer:bind
LOCATION: 12029..12050
OTHER INFORMATION: 17-42.pu
NAME/KEY: primer:bind
LOCATION: 12581..12603
OTHER INFORMATION: 17-42.rp complement
NAME/KEY: primer:bind
LOCATION: 14992..15012
OTHER INFORMATION: 17-41.pu
NAME/KEY: primer:bind
LOCATION: 15460..15482
OTHER INFORMATION: 17-41.rp complement
NAME/KEY: primer:bind
LOCATION: 42070..42090
OTHER INFORMATION: 20-841.pu
NAME/KEY: primer:bind
LOCATION: 42572..42591
OTHER INFORMATION: 20-841.rp complement
NAME/KEY: primer:bind
LOCATION: 45328..45347
OTHER INFORMATION: 20-842.pu
NAME/KEY: primer:bind
LOCATION: 45863..45883
OTHER INFORMATION: 20-842.rp complement
NAME/KEY: primer:bind
LOCATION: 76644..76664
OTHER INFORMATION: 20-853.pu
NAME/KEY: primer:bind
LOCATION: 77166..77185
OTHER INFORMATION: 20-853.rp complement
NAME/KEY: primer:bind
LOCATION: 1220..1238
OTHER INFORMATION: 20-828-311.mis
NAME/KEY: primer:bind
LOCATION: 1240..1258
OTHER INFORMATION: 20-828-311.mis complement
NAME/KEY: primer:bind
LOCATION: 12328..12346
OTHER INFORMATION: 17-42-319.mis
NAME/KEY: primer:bind
LOCATION: 12348..12366
OTHER INFORMATION: 17-42-319.mis complement
NAME/KEY: primer:bind
LOCATION: 15222..15240
OTHER INFORMATION: 17-41-250.mis
NAME/KEY: primer:bind
LOCATION: 15242..15260
OTHER INFORMATION: 17-41-250.mis complement
NAME/KEY: primer:bind
LOCATION: 42199..42217
OTHER INFORMATION: 20-841-149.mis
NAME/KEY: primer:bind
LOCATION: 42219..42237
OTHER INFORMATION: 20-841-149.mis complement
NAME/KEY: primer:bind
LOCATION: 45423..45441
OTHER INFORMATION: 20-842-115.mis
NAME/KEY: primer:bind
LOCATION: 45443..45461
OTHER INFORMATION: 20-842-115.mis complement
```

```
NAME/KEY: primer:bind
LOCATION: 77039..77057
OTHER INFORMATION: 20-853-415.mis
NAME/KEY: primer:bind
LOCATION: 77059..77077
OTHER INFORMATION: 20-853-415.mis complement
NAME/KEY: misc:binding
LOCATION: 1227..1251
OTHER INFORMATION: 20-828-311.probe
NAME/KEY: misc:binding
LOCATION: 12335..12359
OTHER INFORMATION: 17-42-319.probe
NAME/KEY: misc:binding
LOCATION: 15229..15253
OTHER INFORMATION: 17-41-250.probe
NAME/KEY: misc:binding
LOCATION: 42206..42230
OTHER INFORMATION: 20-841-149.probe
NAME/KEY: misc:binding
LOCATION: 45430..45454
OTHER INFORMATION: 20-842-115.probe
NAME/KEY: misc:binding
LOCATION: 77046..77070
OTHER INFORMATION: 20-853-415.probe
US-09-750-590-1
```

```
Query Match 21.6%; Score 26.6; DB 4; Length 81001;
Best Local Similarity 66.7%; Pred. No. 11;
Matches 38; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
```

```
Qy 6 GAAAAAGATTGTTCTAGTGTGGAAGGCCCAACACACTGTCTTTCAGTGGATT 62
Db 77006 GAAAGGACTTTGTTGGGTGGCAGGAGTAACACTGTCTTTAGAAAGAYCGTT 77062
```

RESULT 5

```
US-09-252-991A-778
Sequence 778, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252.991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 778
LENGTH: 786
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-778
```

```
Query Match 21.5%; Score 26.4; DB 4; Length 786;
Best Local Similarity 57.1%; Pred. No. 2.1;
Matches 48; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
```

```
Qy 15 TTGTTCTAGTGTGCGAAAGGCCCAACACTGTCTTTCAGTGGATTAGTTGTACAGA 74
Db 670 TTGATCAGTGTGTGGAGGTGCCCATCTTCTCTTCTCTTCGCTAGGGGTGACGAA 729
Qy 75 ACGGCGTTAGCACTAGCGTTGAC 98
Db 730 AGGGCCGTGGCCCTTTCCTTTAC 753
```

RESULT 6

```
US-09-252-991A-638/c
Sequence 638, Application US/09252991A
```

Db 843 NGNRAATGYGNTG 857

RESULT 8

US-09-706-968-6

; Sequence 6, Application US/09706968

; Patent No. 6528050

; GENERAL INFORMATION:

; APPLICANT: Gao, Zeren

; APPLICANT: Hart, Charles E.

; APPLICANT: Piddington, Christopher S.

; APPLICANT: Sheppard, Paul O.

; APPLICANT: Shoemaker, Kimberly E.

; APPLICANT: Gilbertson, Debra G.

; APPLICANT: West, James W.

; TITLE OF INVENTION: GROWTH FACTOR HOMOLOG ZVEGF3

; FILE REFERENCE: 98-60C1

; CURRENT APPLICATION NUMBER: US/09/706,968

; CURRENT FILING DATE: 2000-11-06

; PRIOR APPLICATION NUMBER: US/09/541,752

; PRIOR FILING DATE: 2000-03-31

; NUMBER OF SEQ ID NOS: 50

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 6

; LENGTH: 1035

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: degenerate sequence derived from SEQ ID NOS: 1 and

; OTHER INFORMATION: 2

; NAME/KEY: misc feature

; LOCATION: (1)_(1035)

; OTHER INFORMATION: n = A,T,C or G

US-09-706-968-6

Query Match 21.5%; Score 26.4; DB 4; Length 1035;

Best Local Similarity 40.0%; Pred. No. 2.3;

Matches 30; Conservative 13; Mismatches 32; Indels 0; Gaps 0;

QY 22 AGTGGTCGAAGGCCCAACACTGTGTTCTGCCAGTGAGTTAGTTGTACAG 81

Db 783 RGARYTNAARMGNACNGAYACNATHTTTGGCCNGGNTGYTNTGTNAARMGNTGYGG 842

QY 82 TAGCACTAGCGCTTG 96

Db 843 NGNRAATGYGNTG 857

RESULT 9

US-09-328-352-2269

; Sequence 2269, Application US/09328352

; Patent No. 6562958

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC99-03PA

; CURRENT APPLICATION NUMBER: US/09/328,352

; CURRENT FILING DATE: 1999-06-04

; NUMBER OF SEQ ID NOS: 8252

; SEQ ID NO 2269

; LENGTH: 1572

; TYPE: DNA

; ORGANISM: Acinetobacter baumannii

US-09-328-352-2269

Query Match 21.5%; Score 26.4; DB 4; Length 1572;

Best Local Similarity 52.8%; Pred. No. 2.8;

Matches 57; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 14 TTGTGTTAGTGGTCGAAGGCCCAACACTGTGTTCTGCCAGTGAGTTAGTTGTACAG 73

Db 1342 TTGTGACAGATTGGTTTACTTCCAAACACTGACTTCTTAAAGACAGTGGAGTTGAGTTA 1401

Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 638

; LENGTH: 900

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-638

Query Match 21.5%; Score 26.4; DB 4; Length 900;

Best Local Similarity 57.1%; Pred. No. 2.2;

Matches 48; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 15 TTGTTCTAGTGGTCGAAGGCCCAACACTGTGTTCTGCCAGTGAGTTAGTTGTACAGA 74

Db 159 TTGATCAGGTGGTGGAGGGTGCCATCTTGTCTGTCTCTCTTCCTCGCTAGGGGTGACGAA 100

QY 75 ACGCGCTTAGCACTAGCGCTTGAC 98

Db 99 AGGCGGTGGCCCTTCCCTTTAC 76

RESULT 7

US-09-457-066-6

; Sequence 6, Application US/09457066

; Patent No. 6432673

; GENERAL INFORMATION:

; APPLICANT: Gao, Zeren

; APPLICANT: Hart, Charles E.

; APPLICANT: Piddington, Christopher S.

; APPLICANT: Sheppard, Paul O.

; APPLICANT: Shoemaker, Kimberly E.

; APPLICANT: Gilbertson, Debra G.

; APPLICANT: West, James W.

; TITLE OF INVENTION: GROWTH FACTOR HOMOLOG ZVEGF3

; FILE REFERENCE: 98-60

; CURRENT APPLICATION NUMBER: US/09/457,066

; CURRENT FILING DATE: 1999-12-07

; NUMBER OF SEQ ID NOS: 50

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 6

; LENGTH: 1035

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: degenerate sequence derived from SEQ ID NOS: 1 and

; OTHER INFORMATION: 2

; NAME/KEY: misc feature

; LOCATION: (1)_(1035)

; OTHER INFORMATION: n = A,T,C or G

US-09-457-066-6

Query Match 21.5%; Score 26.4; DB 4; Length 1035;

Best Local Similarity 40.0%; Pred. No. 2.3;

Matches 30; Conservative 13; Mismatches 32; Indels 0; Gaps 0;

QY 22 AGTGGTCGAAGGCCCAACACTGTGTTCTGCCAGTGAGTTAGTTGTACAGCGCGT 81

Db 783 RGARYTNAARMGNACNGAYACNATHTTTGGCCNGGNTGYTNTGTNAARMGNTGYGG 842

QY 82 TAGCACTAGCGCTTG 96

Db 843 NGNRAATGYGNTG 857

RESULT 9

US-09-328-352-2269

; Sequence 2269, Application US/09328352

; Patent No. 6562958

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC99-03PA

; CURRENT APPLICATION NUMBER: US/09/328,352

; CURRENT FILING DATE: 1999-06-04

; NUMBER OF SEQ ID NOS: 8252

; SEQ ID NO 2269

; LENGTH: 1572

; TYPE: DNA

; ORGANISM: Acinetobacter baumannii

US-09-328-352-2269

Query Match 21.5%; Score 26.4; DB 4; Length 1572;

Best Local Similarity 52.8%; Pred. No. 2.8;

Matches 57; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 14 TTGTGTTAGTGGTCGAAGGCCCAACACTGTGTTCTGCCAGTGAGTTAGTTGTACAG 73

Db 1342 TTGTGACAGATTGGTTTACTTCCAAACACTGACTTCTTAAAGACAGTGGAGTTGAGTTA 1401


```
; 74 AACGGCTTAGCTAGCTGGTTGACAGAACCTCAGACACCAAGGTA 121
; 1402 ACCAACCGTGGCGAGATCATTTTAATGACCGCAACGAAACCAATGTA 1449

RESULT 10
; US-09-214-808-1
; Sequence 1, Application US/09214808A
; Patent No. 6475793
; GENERAL INFORMATION:
; APPLICANT: Rosenthal, Andre
; APPLICANT: Freiberg, Christoph
; APPLICANT: Perret, Xavier Philippe
; APPLICANT: Broughton, William John
; TITLE OF INVENTION: Genomic Sequence of Rhizobium SP. NGR 234 Symbiotic
; Patent No. 6475793
; TITLE OF INVENTION: Plasmid
; FILE REFERENCE: CARP0068
; CURRENT APPLICATION NUMBER: US/09/214,808A
; CURRENT FILING DATE: 1999-06-22
; PRIOR APPLICATION NUMBER: PCT/IB97/00950
; PRIOR FILING DATE: 1997-07-10
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 536165
; TYPE: DNA
; ORGANISM: Rhizobium
; US-09-214-808-1

Query Match 21.5%; Score 26.4; DB 4; Length 536165;
Best Local Similarity 54.0%; Pred. No. 27; Gaps 0;
Matches 54; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 17 GTTCTAGTGTGAAAGGCCCAACTGTGTCTTGTGCCAGTGAGTGTAGTGTGTACAGAAC 76
Db 392324 GATGGAGTGTTAGTAGACCCATCACCCTGGATTCCCAAGAGATATCTCGGCTCAAG 392383

QY 77 GCGTTAGCTAGCTAGCGTTGACAGAACCTCAGACACCAA 116
Db 392384 AGCGATTCAACCGCGCACTAGGCGCAACTCAGAGAGACAA 392423

RESULT 11
; US-08-462-080B-1
; Sequence 1, Application US/08462080B
; Patent No. 5997913
; GENERAL INFORMATION:
; APPLICANT: Fowler, Timothy
; APPLICANT: Shoemaker, Sharon
; APPLICANT: Barnett, Christopher C.
; TITLE OF INVENTION: Saccharification of Cellulose by Cloning and
; TITLE OF INVENTION: Amplification of the Beta-glucosidase Gene of Trichoderma Reesei
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International, Inc.
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: Ca
; COUNTRY: U.S.A.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,080B
; FILING DATE: 05-JUN-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/248,586
```

```
; FILING DATE: 24-MAY-1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/807,028
; FILING DATE: 10-DEC-1991
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/625,140
; FILING DATE: 10-DEC-1990
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Stone, Christopher L.
; REGISTRATION NUMBER: 35,696
; REFERENCE/DOCKET NUMBER: GC78D3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-846-7555
; TELEFAX: 650-845-6504
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3033 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Trichoderma reesei
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(311...375, 446...2205, 2270...2679)
; FEATURE:
; NAME/KEY: intron
; LOCATION: 376...445
; FEATURE:
; NAME/KEY: intron
; LOCATION: 2206...2269
; US-08-462-080B-1

Query Match 20.8%; Score 25.6; DB 2; Length 3033;
Best Local Similarity 55.7%; Pred. No. 7.1;
Matches 49; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 12 AGTTGTTCTAGTGTGCGAAAGGCCCAACTGTGTTCGCCAGTGAGTGTGTAC 71
Db 260 AGTTCTCTTACCAAGTCTTGACCAAGACCATCTGTGTAGCCCAATCAGAAATGCGTTAC 319

QY 72 AGAACGCGTTAGCACTAGCGCTTGACA 99
Db 320 CGAACAGACGCTCGCTGGCACTTGCCA 347

RESULT 12
; US-08-462-090-1
; Sequence 1, Application US/08462090
; Patent No. 6022725
; GENERAL INFORMATION:
; APPLICANT: Fowler, Timothy
; APPLICANT: Barnett, Christopher C.
; APPLICANT: Shoemaker, Sharon
; TITLE OF INVENTION: Saccharification of Cellulose by Cloning
; TITLE OF INVENTION: and Amplification of the Beta-glucosidase Gene of
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Building, 699 Prince St.
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22313-1404
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
```

```

,
, COMPUTER: IBM PC compatible
,
, OPERATING SYSTEM: PC-DOS/MS-DOS
,
, SOFTWARE: PatentIn Release #1.0, Version #1.25
,
, CURRENT APPLICATION DATA:
,
, APPLICATION NUMBER: US/08/463,461
,
, FILING DATE: 05-JUN-1995
,
, CLASSIFICATION: 435
,
, ATTORNEY/AGENT INFORMATION:
,
, NAME: Christopher L. Stone
,
, REGISTRATION NUMBER: 35,696
,
, REFERENCE/DOCKET NUMBER: G78D4
,
, TELECOMMUNICATION INFORMATION:
,
, TELEPHONE: 415-846-7555
,
, TELEFAX: 415-845-6504
,
, INFORMATION FOR SEQ ID NO: 1:
,
, SEQUENCE CHARACTERISTICS:
,
, LENGTH: 3033 base pairs
,
, TYPE: nucleic acid
,
, STRANDEDNESS: single
,
, TOPOLOGY: linear
,
, MOLECULE TYPE: DNA (genomic)
,
, HYPOTHETICAL: NO
,
, ANTI-SENSE: NO
,
, ORIGINAL SOURCE:
,
, ORGANISM: Trichoderma reesei
,
, FEATURE:
,
, NAME/KEY: CDS
,
, LOCATION: join(311..375, 446..2205, 2270..2679)
,
, FEATURE:
,
, NAME/KEY: intron
,
, LOCATION: 376..445
,
, FEATURE:
,
, NAME/KEY: intron
,
, LOCATION: 2206..2269
,
, US-08-463-461-1

```

Query Match	20.8%;	Score 25.6;	DB 3;	Length 3033;
Best Local Similarity	55.7%;	Pred. No. 7.1;	39;	Gaps 0;
Matches	49;	Conservative	0;	Mismatches
Qy	12	AGTTCTTCTAGTGGTGAAGGCCCAACACTCTGTTTTCGACGTAGTTCAGTTGTAC	71	
Db	260	AGTTCTTCTTACCAGTCTTCACCAAGACCAATCTGTGACCCCAATCAGAAATGCGTTAC	319	
Qy	72	AGAACGCGGTTAGCACTAGCGCTTGACA	99	
Db	320	CGAACAGCACTGCGGTGGCACTTGCA	347	

```

RESULT 14
US-08-916-421B-1/c
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OP INVENTION: Complete Genome Sequen
; Patent No. 6503729
; FILE OF INVENTION: jannaschii
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: t equals a, t.c, or g

```

```
NAME/KEY: misc feature
LOCATION: (28257)..(28258)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84773)..(84773)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84808)..(84808)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84812)..(84812)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98120)..(98120)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98159)..(98159)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98239)..(98239)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98266)..(98266)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98343)..(98343)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (103998)..(103998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (148948)..(148948)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (163385)..(163385)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191989)..(191989)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191995)..(191995)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (231980)..(231980)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234187)..(234187)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234220)..(234220)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234814)..(234814)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309398)..(309398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309418)..(309418)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312837)..(312837)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312993)..(312993)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (319226)..(319226)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (559167)..(559167)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (559241)..(559241)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (600992)..(600992)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (622708)..(622708)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (657081)..(657081)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (657203)..(657203)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (674435)..(674435)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (682442)..(682442)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (713652)..(713652)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (741684)..(741684)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779455)..(779455)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779676)..(779676)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (855539)..(855539)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (871619)..(871619)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1084830)..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1096846)..(1096846)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1119881)..(1119881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1130881)..(1130881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1310988)..(1310988)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1313224)..(1313224)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1349473)..(1349473)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1349491)..(1349491)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1470091)..(1470091)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1569020)..(1569020)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1602912)..(1602912)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1603734)..(1603734)
```

OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1637998)..(1637998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1667854)..(1667854)
OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

Query Match 20.7%; Score 25.4; DB 4; Length 1664976;
Best Local Similarity 58.7%; Pred. No. 84;
Matches 44; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
QY 10 AAAGTTTGTCTAGTGGTCGAAGGCCCAACACTGTGTTCTTGCCAGTGAGTTGTTGTAAGAACCGCGT 81
DB 1601936 AATATTTATGCTGTGTGTAAGGCCCAACACTGTGTTCTTGCCAGTGAGTTGTTGTAAGAACCGCGT 47861
QY 70 ACAGAACGGCGGTAG 84
DB 1601876 CTTTAACTGCGAGAG 1601862

RESULT 15
US-08-614-770A-1/c
Sequence 1, Application US/08614770A
Patent No. 5773267
GENERAL INFORMATION:
APPLICANT: WILLIAM R. JACOBS AND GRAHAM F. HATFULL
TITLE OF INVENTION: D29 SHUTTLE PHASMIDS AND USES THEREOF
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
STREET: 90 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH 1.44 Mb STORAGE
MEDIUM TYPE: DISKETTE
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/614,770A
FILING DATE: MARCH 7, 1996
ATTORNEY/AGENT INFORMATION:
NAME: ELIZABETH A. BOGOSIAN
REGISTRATION NUMBER: 39,911
REFERENCE/DOCKET NUMBER: 96700/402
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-5995
TELEFAX: (212) 286-0854 or 286-0082
TELEX: TWX 710-581-4766
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 49272
TYPE: NUCLEIC ACID
STRANDEDNESS: DOUBLE
TOPOLOGY: LINEAR
MOLECULE TYPE:
DESCRIPTION: OLIGONUCLEOTIDE
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: MYCOBACTERIOPHAGE
INDIVIDUAL ISOLATE: D29
US-08-614-770A-1

Query Match 20.5%; Score 25.2; DB 1; Length 49272;
Best Local Similarity 55.8%; Pred. No. 30;
Matches 48; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 22 AGTGGTCGAAAGGCCCAACACTGTGTTCTTGCCAGTGAGTTGTTGTAAGAACCGCGT 81
DB 47920 AGTGGTCGAAAGGCCCAACACTGTGTTCTTGCCAGTGAGTTGTTGTAAGAACCGCGT 47861
QY 82 TAGCACTAGCGGCTTGACAGAACCTCA 107
DB 47860 CATCAGCATCACTGGCGGACACCGTCA 47835

Search completed: November 23, 2003, 15:53:39
Job time : 19.3791 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 23, 2003, 14:02:08 ; Search time 43.5904 Seconds
(without alignments)
9221.962 Million cell updates/sec

Title: US-09-717-321A-1

Perfect score: 123

Sequence: 1 caattgaaaaagtgtgttc.....ctcacagacccaaggtacc 123

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2169961 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

```
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US05_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	79.6	64.7	1232	10	US-09-880-107-1632
C 2	79.6	64.7	1644	14	US-10-153-668-48
C 3	79.6	64.7	1646	14	US-10-153-668-46
C 4	79.6	64.7	2051	9	US-09-925-302-255
C 5	79.6	64.7	2941	14	US-10-198-846-9874
C 6	57	46.3	353	10	US-09-960-352-11432
C 7	38.2	31.1	3740	11	US-09-764-891-9984
C 8	33.4	27.2	2313	14	US-10-198-846-12919
C 9	31.4	25.5	477	14	US-10-066-543-3022
C 10	31.4	25.5	549	14	US-10-066-543-2878
C 11	31.4	25.5	567	14	US-10-066-543-2964
C 12	31.4	25.5	602	10	US-09-878-178-1654
C 13	31.4	25.5	602	13	US-10-046-935-1654
C 14	31.4	25.5	602	14	US-10-146-502-1654
C 15	31.4	25.5	605	11	US-09-871-161-483
C 16	29.4	23.9	509	12	US-10-027-632-323187

```
c 17 29.4 23.9 509 13 US-10-027-632-323187
c 18 28.8 23.4 202001 9 US-09-734-674-3
c 19 28.8 23.1 202001 14 US-10-274-990-3
c 20 28.4 23.1 594 14 US-10-066-543-2864
c 21 27.4 22.3 412 10 US-09-867-701-3218
c 22 27.4 22.3 412 10 US-03-880-107-712
c 23 27.4 22.3 412 10 US-03-954-531-88
c 24 27.4 22.3 412 10 US-03-954-531-288
c 25 27.4 22.3 412 10 US-09-954-531-504
c 26 27.4 22.3 486 11 US-09-918-995-25261
c 27 27.4 22.3 1658 14 US-10-171-311-237
c 28 27.2 22.1 9965 12 US-10-311-455-1499
c 29 27 22.0 6693 13 US-10-071-766-135
c 30 26.8 21.8 746 12 US-10-027-632-150772
c 31 26.8 21.8 746 12 US-10-027-632-150773
c 32 26.8 21.8 746 13 US-10-027-632-150772
c 33 26.8 21.8 746 13 US-10-027-632-150773
c 34 26.8 21.8 1617 10 US-09-917-800A-1699
c 35 26.8 21.8 1617 12 US-10-205-194-175
c 36 26.8 21.8 8409 11 US-09-808-602-79
c 37 26.8 21.8 8409 11 US-09-800-198-67
c 38 26.6 21.6 631 12 US-10-027-632-291355
c 39 26.6 21.6 631 12 US-10-027-632-291356
c 40 26.6 21.6 631 13 US-10-027-632-291355
c 41 26.6 21.6 631 13 US-10-027-632-291356
c 42 26.6 21.6 1011 12 US-10-027-632-323948
c 43 26.6 21.6 1011 13 US-10-027-632-323948
c 44 26.6 21.6 81001 10 US-09-751-877-1
c 45 26.6 21.6 81001 11 US-09-842-364-1
```

ALIGNMENTS

RESULT 1

```
US-09-880-107-1632/c
; Sequence 1632, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1632
; LENGTH: 1232
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 D25274
US-09-880-107-1632
```

Query Match 64.7%; Score 79.6; DB 10; Length 1232;
Best Local Similarity 83.1%; Pred. No. 2.8e+20;
Matches 103; Conservative 0; Mismatches 19; Indels 2; Gaps 1;

```
Qy 1 CAATTGAAAAGTTCTTCTAGTGGTCGAAAGCCCAACTGTGTTCTTGCAGTAG 60
    |||||
Db 1135 CAATTAAAAAATCTCTTAGTGGCTGAAGGTCCTCCACGCTGTATTCTGCCAGTAG 1076
    |||||

Qy 61 TTAGTTGTACAGAACCGGCTTAGCACTAGCGC--TTGACAGAACCTCAGACCCCAAG 118
    |||||
Db 1075 TTAGTTGTACAGAACATCGTCAGCACTAGCAGATTTCAGAACCTCAGACCCCAAG 1016
    |||||
```

```
QY 119 GTAC 122
Db 1015 GAAC 1012

RESULT 2
US-10-153-668-48/c
; Sequence 48, Application US/10153668
; Publication No. US20030092616A1
; GENERAL INFORMATION:
; APPLICANT: HONDA, Goichi
; APPLICANT: MATSUDA, Akio
; APPLICANT: MURAMATSU, Shuji
; APPLICANT: ISHIZAWA, Kenya
; TITLE OF INVENTION: STAT6 Activating Gene
; FILE REFERENCE: 1254-0207P
; CURRENT APPLICATION NUMBER: US/10/153,668
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: US 60/293,172
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/316,031
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/328,403
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: JP 2001-157043
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: JP 2001-260681
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: JP 2001-313175
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 488
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 48
; LENGTH: 1644
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (109)..(447)
; US-10-153-668-48

Query Match 64.7%; Score 79.6; DB 14; Length 1644;
Best Local Similarity 83.1%; Pred. No. 3.1e-20;
Matches 103; Conservative 0; Mismatches 19; Indels 2; Gaps 1;

QY 1 CAATTGAAAAAGTTTCTTAGTGGTCGAAAGGCCCAACACTGTGTCTTGGCAGTGAG 60
|||||
Db 1550 CAATTTAAAAAATCTGTGTAGTGGTGAAGGTCCTCCACGCTGATTCTCGCCAGTGAG 1491
|||||

QY 61 TTAGTTGTACAGAACGGCGTTAGCACTAGCGC--TTGACAGAACCTTCAGACCCCAAG 118
|||||
Db 1490 TTAAGTTGTACAGAACATCGTCAGCACTAGCAGATTACAGAACCTTCAGACCCCAAG 1431
|||||

QY 119 GTAC 122
Db 1430 GAAC 1427

RESULT 4
US-09-925-302-255/c
; Sequence 255, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: FA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 255
; LENGTH: 2051
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (50)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (68)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (2027)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (2046)

Query Match 64.7%; Score 79.6; DB 14; Length 1644;
Best Local Similarity 83.1%; Pred. No. 3.1e-20;
Matches 103; Conservative 0; Mismatches 19; Indels 2; Gaps 1;

QY 1 CAATTGAAAAAGTTTCTTAGTGGTCGAAAGGCCCAACACTGTGTCTTGGCAGTGAG 60
|||||
Db 1548 CAATTTAAAAAATCTGTGTAGTGGTGAAGGTCCTCCACGCTGATTCTCGCCAGTGAG 1489
|||||

QY 61 TTAGTTGTACAGAACGGCGTTAGCACTAGCGC--TTGACAGAACCTTCAGACCCCAAG 118
|||||
Db 1488 TTAAGTTGTACAGAACATCGTCAGCACTAGCAGATTACAGAACCTTCAGACCCCAAG 1429
|||||

QY 119 GTAC 122
Db 1428 GAAC 1425

RESULT 3
US-10-153-668-46/c
; Sequence 46, Application US/10153668
; Publication No. US20030092616A1
; GENERAL INFORMATION:
; APPLICANT: HONDA, Goichi
; APPLICANT: MATSUDA, Akio
; APPLICANT: MURAMATSU, Shuji
; APPLICANT: ISHIZAWA, Kenya
; TITLE OF INVENTION: STAT6 Activating Gene
; FILE REFERENCE: 1254-0207P
; CURRENT APPLICATION NUMBER: US/10/153,668
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: US 60/293,172
; PRIOR FILING DATE: 2001-05-25
```

```
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-302-255
Query Match      64.7%; Score 79.6; DB 9; Length 2051;
Best Local Similarity 83.1%; Pred. No. 3.4e-20;
Matches 103; Conservative 0; Mismatches 19; Indels 2; Gaps 1;

QY 1 CAATTGAAAAGTTTCTGTAGTGGTTCGAAAGGCCCAACACTGTGTTCTTGCCAGTGAG 60
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1996 CAATTGAAAAGTTTCTGTAGTGGTTCGAAAGGCCCAACACTGTGTTCTTGCCAGTGAG 1837

QY 61 TTAGTTGTACAGAACGGCGTTAGCACTAGCGC--TTGACAGAAACCTCACAGACCCAAAG 118
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1936 TTAGTTGTACAGAACATCGTCAGCACTAGCACTAGCACTTACAGAACTTACAGACCCAAAG 1777

QY 119 GTAC 122
Db ||||
1776 GAAC 1773

RESULT 5
US-10-198-846-9874/c
; Sequence 9874, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 9874
; LENGTH: 2941
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-198-846-9874

Query Match      64.7%; Score 79.6; DB 14; Length 2941;
Best Local Similarity 83.1%; Pred. No. 3.9e-20;
Matches 103; Conservative 0; Mismatches 19; Indels 2; Gaps 1;

QY 1 CAATTGAAAAGTTTCTGTAGTGGTTCGAAAGGCCCAACACTGTGTTCTTGCCAGTGAG 60
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2336 CAATTGAAAAGTTTCTGTAGTGGTTCGAAAGGCCCAACACTGTGTTCTTGCCAGTGAG 2277

QY 61 TTAGTTGTACAGAACGGCGTTAGCACTAGCGC--TTGACAGAAACCTCACAGACCCAAAG 118
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2276 TTAGTTGTACAGAACATCGTCAGCACTAGCACTAGCACTTACAGAACTTACAGACCCAAAG 2217

QY 119 GTAC 122
Db ||||
2216 GAAC 2213

RESULT 6
US-09-960-352-11432
; Sequence 11432, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
```

```
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 11432
; LENGTH: 353
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 49-LIB3058-009-Q1-K1-E2
US-09-960-352-11432

Query Match      46.3%; Score 57; DB 10; Length 353;
Best Local Similarity 82.8%; Pred. No. 9.9e-12;
Matches 77; Conservative 0; Mismatches 15; Indels 1; Gaps 1;

QY 1 CAATTGAAAAGTTTCTGTAGTGGTTCGAAAGGCCCAACACTGTGTTCTTGCCAGTGAG 60
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
155 CAATTGAAAAGTTTCTGTAGTGGTTCGAAAGGCCCAACACTGTGTTCTTGCCAGTG-G 213

QY 61 TTAGTTGTACAGAACGGCGTTAGCACTAGCGC 93
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
214 GTAAGTTGTACAGAACTTCGTTAGCAGGAC 246

RESULT 7
US-09-764-891-9984/c
; Sequence 9984, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9984
; LENGTH: 3740
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-9984

Query Match      31.1%; Score 38.2; DB 11; Length 3740;
Best Local Similarity 78.0%; Pred. No. 0.00048;
Matches 46; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 8 AAAAAAGTTTCTGTAGTGGTTCGAAAGGCCCAACACTGTGTTCTTGCCAGTGAGTTAGGT 66
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2725 AAAAAATTTGTTCTAGTTGTTGAAAGGCCCAAGCTTATTCITGCGCAGTCTTAAGGT 2667

RESULT 8
US-10-198-846-12919/c
; Sequence 12919, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 12919
; LENGTH: 2313
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-198-846-12919

Query Match      27.2%; Score 33.4; DB 14; Length 2313;
Best Local Similarity 86.0%; Pred. No. 0.029;
Matches 49; Conservative 0; Mismatches 6; Indels 2; Gaps 1;

QY 68 GTACAGACGGCGTTAGCACTAGCGC--TTGACAGAACTTCACAGACCCCAAGGTAC 122
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2296 GTACAGACATCGTCAGCACTAGCAGAGTTTACAGAACTTCACAGACCCCAAGGAAC 2240

RESULT 9
US-10-066-543-3022
; Sequence 3022, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indirias, Carol Yoseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Margarita
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.563
; CURRENT APPLICATION NUMBER: US/10/066,543
; CURRENT FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3022
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-066-543-3022

Query Match      25.5%; Score 31.4; DB 14; Length 477;
Best Local Similarity 85.5%; Pred. No. 0.093;
Matches 47; Conservative 0; Mismatches 6; Indels 2; Gaps 1;

QY 70 ACAGAACGGCGTTAGCACTAGCGC--TTGACAGAACTTCACAGACCCCAAGGTAC 122
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 ACAGAACATCGTCAGCACTAGCAGAGTTTACAGAACTTCACAGACCCCAAGGAAC 55

RESULT 10
US-10-066-543-2878
; Sequence 2878, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indirias, Carol Yoseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Margarita
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.563
; CURRENT APPLICATION NUMBER: US/10/066,543
; CURRENT FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 3417
```

```
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2878
; LENGTH: 549
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 526
; OTHER INFORMATION: n = A,T,C or G
US-10-066-543-2878

Query Match      25.5%; Score 31.4; DB 14; Length 549;
Best Local Similarity 85.5%; Pred. No. 0.098;
Matches 47; Conservative 0; Mismatches 6; Indels 2; Gaps 1;

QY 70 ACAGAACGGCGTTAGCACTAGCGC--TTGACAGAACTTCACAGACCCCAAGGTAC 122
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 ACAGAACATCGTCAGCACTAGCAGAGTTTACAGAACTTCACAGACCCCAAGGAAC 55

RESULT 11
US-10-066-543-2964
; Sequence 2964, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indirias, Carol Yoseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Margarita
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.563
; CURRENT APPLICATION NUMBER: US/10/066,543
; CURRENT FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2964
; LENGTH: 567
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 527
; OTHER INFORMATION: n = A,T,C or G
US-10-066-543-2964

Query Match      25.5%; Score 31.4; DB 14; Length 567;
Best Local Similarity 85.5%; Pred. No. 0.099;
Matches 47; Conservative 0; Mismatches 6; Indels 2; Gaps 1;

QY 70 ACAGAACGGCGTTAGCACTAGCGC--TTGACAGAACTTCACAGACCCCAAGGTAC 122
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 ACAGAACATCGTCAGCACTAGCAGAGTTTACAGAACTTCACAGACCCCAAGGAAC 55

RESULT 12
US-09-878-178-1654
; Sequence 1654, Application US/09878178
; Patent No. US20020177552A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.527
```



```

/ APPLICANT: Jiang, Yuqiu
/ APPLICANT: Harlocker, Susan L.
/ APPLICANT: Secrist, Heather
/ APPLICANT: Wang, Aijun
/ APPLICANT: Stolk, John A.
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
/ OF COLON CANCER
/ FILE REFERENCE: 210121.527C2
/ CURRENT APPLICATION NUMBER: US/10/146.502

```

Search completed: November 23, 2003, 19:05:39
Job time : 47.5904 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: November 23, 2003, 09:18:57 ; Search time 3810.98 Seconds
(without alignments)
10917.162 Million cell updates/sec
Title: US-09-717-321A-15
Perfect score: 1017
Sequence: 1 cccctattctgtcagatt.....cctttgggtctgtgaggttc 1017
Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 2888711 seqs, 20454813386 residues
Total number of hits satisfying chosen parameters: 5777422
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1017	100.0	1017	6	AX163751	AX163751 Sequence
c	1017	100.0	1266	6	AX163738	AX163738 Sequence
3	1013.8	99.7	217700	2	AC106124	AC106124 Rattus no
4	679.2	66.8	2281	10	BC051053	BC051053 Mus muscu
5	677.6	66.6	2319	10	BC003828	BC003828 Mus muscu
6	677.6	66.6	269081	2	AC068493	AC068493 Mus muscu
7	403.2	39.6	185245	2	AC102775	AC102775 Mus muscu
8	403.2	39.6	202342	10	AL583884	AL583884 Mouse DNA
9	337.2	33.2	179685	10	AC126556	AC126556 Mus muscu
10	290.6	28.6	1022	6	AX163752	AX163752 Sequence
11	290.6	28.6	2302	9	BC050687	BC050687 Homo sapi
12	290.6	28.6	2315	9	AK054993	AK054993 Homo sapi
13	290.6	28.6	28567	9	HSAL32695	HSAL32695 Homo sapi
14	290.6	28.6	212827	9	AC009412	AC009412 Homo sapi
15	285.2	28.0	1232	6	AX408985	AX408985 Sequence
16	285.2	28.0	1232	6	AX104663	AX104663 Homo sapi
c	173	17.0	137625	9	AC104663	AC104663 Homo sapi
17	194.6	19.1	202565	9	HMP02879	D52874 Homo sapien
c	193.2	19.0	605	6	AL354696	AL354696 Human DNA
c	190.2	18.7	192498	2	AX341407	AX341407 Sequence
c	190.2	18.7	230015	2	AC105979	AC105979 Mus muscu
c	173	17.0	137625	9	AC132602	AC132602 Mus muscu
c	169.8	16.7	5544	9	AC104663	AC104663 Homo sapi
c	116.2	11.4	137625	9	AF542527	AF542527 Homo sapi
c	113.8	11.2	87616	2	AC104663	AC104663 Homo sapi
25	113	11.1	348	6	AX163753	AX163753 Sequence
26	112.4	11.1	64781	2	AC102025	AC102025 Mus muscu
c	105.8	10.4	101584	9	CNS01DS5	AL121655 BAC seque
c	105.8	10.4	155943	9	AC012364	AC012364 Homo sapi
c	103.4	10.2	5544	9	AF542527	AF542527 Homo sapi
c	103.2	10.1	110816	9	AC002404	AC002404 Human Chr
c	98.4	9.7	170839	2	AX133467	AX133467 Mus muscu
c	85.6	8.4	495	6	AX341256	AX341256 Sequence
c	84.2	8.3	180303	9	AL672045	AL672045 Human DNA
c	84.2	8.3	201012	2	AC021189	AC021189 Homo sapi
c	76.2	7.5	228121	2	AC133022	AC133022 Rattus no
c	76.2	7.5	239768	2	AC112582	AC112582 Rattus no
c	76.2	7.5	245468	2	AC130746	AC130746 Rattus no
c	67	6.6	594	6	AX385363	AX385363 Sequence
c	62.8	6.2	194371	2	AC135453	AC135453 Rattus no
c	61.4	6.0	378	6	BD030694	BD030694 Sequence
c	56.8	5.6	198	11	G31747	G31747 sWSS2749 Er
c	54.2	5.3	7218	6	I66494	I66494 Sequence 14
c	52.4	5.2	162	6	AX397334	AX397334 Sequence
c	51.4	5.1	90354	9	AP001124	AP001124 Homo sapi
c	51.4	5.1	178089	2	AC010929	AC010929 Homo sapi

ALIGNMENTS

RESULT 1
AX163751
LOCUS AX163751 1017 bp DNA linear PAT 22-JUN-2001
DEFINITION Sequence 15 from Patent WO0138579.
ACCESSION AX163751
VERSION AX163751.1 GI:14544857
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Bukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1
AUTHORS Gould-Rothberg,B.E., Dipippo,V.A., Ramseh,T.M. and Gerwein,R.W.
TITLE Method of identifying toxic agents using nsaid-induced differential


```
QY 481 CATTGTACATGAGTAATCACTCAATAAAGTGTACGGTAAAGCGTTTAAACGGTT 540
Dd |||||
QY 781 CATTGTACATGAGTAATCACTCAATAAAGTGTACGGTAAAGCGTTTAAACGGTT 722
Dd |||||
QY 541 AATTTCCTGTCACAAAGTAGATGACAAATAGGCGGATCTTATCAGTGTCTCTCTGAGCCCC 600
Dd |||||
QY 721 AATTTCCTGTCACAAAGTAGATGACAAATAGGCGGATCTTATCAGTGTCTCTCTGAGCCCC 662
QY 601 CTTTCCCGCTGCTGCTCCCTCCAGATGGGGGCTTGAAGTCATATTTAAACTGCGCATATG 660
Dd |||||
QY 661 CTTTCCCGCTGCTGCTCCCTCCAGATGGGGGCTTGAAGTCATATTTAAACTGCGCATATG 602
QY 661 TCACAGTTGCTAACTTAGCAAGTGTCTTCTTTAGGACCCCTCTTAAACGAGCAATATG 720
Dd |||||
QY 601 TCACAGTTGCTAACTTAGCAAGTGTCTTCTTTAGGACCCCTCTTAAACGAGCAATATG 542
QY 721 TCTGACCTGTACTATAAGATCTTCTGATAATGATTCGGAGATTTTTTTCGTTAGATAGT 780
Dd |||||
QY 541 TCTGACCTGTACTATAAGATCTTCTGATAATGATTCGGAGATTTTTTTCGTTAGATAGT 482
QY 781 AGAAGTGGTTCCTGTTTACCTTCCTTACTCAGTGTAGTGTCTTCCCTTCGTTTTT 840
Dd |||||
QY 481 AGAAGTGGTTCCTGTTTACCTTCCTTACTCAGTGTAGTGTCTTCCCTTCGTTTTT 422
QY 841 CTAGTAACCTGGGTAGAAATCAAGTGTCTGGGCTTTACAGTGTTTTAAACATTTTATAGT 900
Dd |||||
QY 421 CTAGTAACCTGGGTAGAAATCAAGTGTCTGGGCTTTACAGTGTTTTAAACATTTTATAGT 362
QY 901 ATTCTGAAACATCACTGTTTCCAGATACCAACATGTCTCATGTATGATGCGGCC 960
Dd |||||
QY 361 ATTCTGAAACATCACTGTTTCCAGATACCAACATGTCTCATGTATGATGCGGCC 302
QY 961 CTCTAGACCTCACCACCGGACATGCTTCCGTTACCTTGGTCTGTGAGGTTTC 1017
Dd |||||
QY 301 CTCTAGACCTCACCACCGGACATGCTTCCGTTACCTTGGTCTGTGAGGTTTC 245

RESULT 3
AC106124
LOCUS
DEFINITION
Rattus norvegicus clone CH230-119E11, *** SEQUENCING IN PROGRESS
AC106124
AC106124 6 GI:30521557
HTG; HTGS_PRAISE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 217700)
Muzny,D,Marle., Metzker,M, Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Cesar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyie,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Cartoll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gregeorge,P., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpthy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
```

```
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhwa,L., Loulseghe,H., Lozada,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Muniadasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwackelneh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.,
Puzo,M., Quiroz,J., Rachlin,S., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rockey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 217700)
Worley,K.C.
Direct Submission
Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 217700)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 10, 2003 this sequence version replaced gi:25094619.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GHLL
Center clone name: CH230-119E11
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 185725 bases at least Q40
Consensus quality: 190140 bases at least Q30
Consensus quality: 192842 bases at least Q20
Estimated insert size: 194758; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank/draft.data.html).
* NOTE: This is a 'working draft' sequence. It currently
```

* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
1 109797: contig of 109797 bp in length
* 109798: gap of unknown length
* 109898: contig of 57037 bp in length
* 166935: gap of unknown length
* 166935: contig of 57037 bp in length
* 167035: gap of unknown length
* 167035: contig of 9350 bp in length
* 176385: gap of unknown length
* 176385: contig of 29843 bp in length
* 206328: gap of unknown length
* 206328: contig of 1034 bp in length
* 207462: gap of unknown length
* 207462: contig of 1046 bp in length
* 208607: gap of unknown length
* 208607: contig of 1201 bp in length
* 209908: gap of unknown length
* 209908: contig of 1575 bp in length
* 211583: gap of unknown length
* 211583: contig of 1966 bp in length
* 213649: gap of unknown length
* 213649: contig of 1366 bp in length
* 213750: gap of unknown length
* 215116: contig of 2485 bp in length.
* 215216: contig of 2485 bp in length.
Location/Qualifiers
1. .217700
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-119E11"
BASE COUNT 51384 a 45640 c 45954 g 51476 t 23246 others
ORIGIN
```

FEATURES

Source

```
Query Match 99.7%; Score 1013.8; DB 2; Length 217700;
Best Local Similarity 99.8%; Pred. No. 1.7e-290;
Matches 1015; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCCTATTCTGCTCAGATTAAGAAATGCCAAATACCTTGTAATGCTGCTGCTG 60
Db 63447 CCCTATTCTGCTCAGATTAAGAAATGCCAAATACCTTGTAATGCTGCTG 63506

QY 61 CTGAGAACGTAAGCACTAAGCTGTGTGAGAGACTTTGCTCTTAAAGAACTGCAGCTTC 120
Db 63507 CTGAGAACGTAAGCACTAAGCTGTGTGAGAGACTTTGCTCTTAAAGAACTGCAGCTTC 63566

QY 121 TGGGCTCAGGGGTGCAGACCTCCGCTAGCTCCAGACCGGTGTGACACAGCCTCC 180
Db 63567 TGGGCTCAGGGGTGCAGACCTCCGCTAGCTCCAGACCGGTGTGACACAGCCTCC 63626

QY 181 TTAATGACACGCTGCCATGTAAAGCACTGTAATCTATAGCCCATGCTCATTAACGTAAC 240
Db 63627 TTAATGACACGCTGCCATGTAAAGCACTGTAATCTATAGCCCATGCTCATTAACGTAAC 63686

QY 241 TTGTGACTGTACGTACGATGGGTGTAAAGCTCTGCTCTTTGATTTATAGTGGTTCT 300
Db 63687 TTGTGACTGTACGTACGATGGGTGTAAAGCTCTGCTCTTTGATTTATAGTGGTTCT 63746

QY 301 CTAATAATACAGCTGACCGCTCTTCAGGCTTTGAACAGAACTCTGGCTCTGTTGTC 360
Db 63747 CTAATAATACAGCTGACCGCTCTTCAGGCTTTGAACAGAACTCTGGCTCTGTTGTC 63806

QY 361 CTCTAAGAAAGTATTCTGTTCTTAGTGTGGGTGTGGGTGGAGTGTGTGAACACAGA 420
Db 63807 CTCTAAGAAAGTATTCTGTTCTTAGTGTGGGTGTGGGTGGAGTGTGTGAACACAGA 63866

QY 421 CGTCTCAAGAGGAGACAGACAGTATTTTGACTAATATGAAAGTGAAGATTAATTTACACTA 480
Db 63867 CGTCTCAAGAGGAGACAGACAGTATTTTGACTAATATGAAAGTGAAGATTAATTTACACTA 63926
```

```
QY 481 CATTGTACATGAGTAATTCACATGAATAAAGTGTACGGGTAAAGCTTTTAAACGGTT 540
Db 63927 CATTGTACATGAGTAATTCACATGAATAAAGTGTACGGGTAAAGCTTTTAAACGGTT 63986

QY 541 AATTTCTGTCAAAACAGTAGATGACAAATGCGCGATCTTATCAGTGTCTCTCTTGAGCCCC 600
Db 63987 AATTTCTGTCAAAACAGTAGATGACAAATGCGCGATCTTATCAGTGTCTCTCTTGAGCCCC 64046

QY 601 CTTTCCCTCTGTCTGCTCCCTCCAGATGCGGGGTGTAGTCCATATTTAAACTGGCCATCC 660
Db 64047 CTTTCCCTCTGTCTGCTCCCTCCAGATGCGGGGTGTAGTCCATATTTAAACTGGCCATCC 64106

QY 661 TCACAGTGTCTAACTAGCAAGTCTTTCTTTAGACCCCTCTTAAAGCAACAATATG 720
Db 64107 TCACAGTGTCTAACTAGCAAGTCTTTCTTTAGACCCCTCTTAAAGCAACAATATG 64166

QY 721 TCTGACCTGTACTATAAAGATCTTTCTGATAATGCAATCGGAGATTTTTTGTAGATAGT 780
Db 64167 TCTGACCTGTACTATAAAGATCTTTCTGATAATGCAATCGGAGATTTTTTGTAGATAGT 64226

QY 781 AGAAGTGGTTCCTGTTTTCACCTCTTACTCAGCTGACTAGTCCCTTCGTTT 840
Db 64227 AGAAGTGGTTCCTGTTTTCACCTCTTACTCAGCTGACTAGTCCCTTCGTTT 64286

QY 841 CTAGTAACCTGGGTGTAGAAATCAGCTGCTGGGCTTTACAGTCTTTTAAACTATTTAGAT 900
Db 64287 CTAGTAACCTGGGTGTAGAAATCAGCTGCTGGGCTTTACAGTCTTTTAAACTATTTAGAT 64346

QY 901 ATTCTGAAACATCACTGTCTGTCAGAGTACCAACACTGTGATGATGATCCGCCCC 960
Db 64347 ATTCTGAAACATCACTGTCTGTCAGAGTACCAACACTGTGATGATGATCCGCCCC 64406

QY 961 CTCTAGACCTCACCACGGGACACATGCTTCGGGTACCTTTGGTCTGTGAGGTTTC 1017
Db 64407 CTCTAGACCTCACCACGGGACACATGCTTCGGGTACCTTTGGTCTGTGAGGTTTC 64463
```

```
RESULT 4
LOCUS BC051053 2281 bp mRNA linear ROD 14-APR-2003
DEFINITION Mus musculus, clone MGC:58966 IMAGE:5038182, mRNA, complete cds.
ACCESSION BC051053
VERSION BC051053.1 GI:29835221
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 2281)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (11-APR-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgi.nci.nih.gov
Contact: nisc.mgc@nih.gov
Email: cgabps-femail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc.mgc@nih.gov
Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
Maduro,Q.J., Masello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
Tsurgonov,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
```

Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAK Plate: 108 Row: f Column: 9
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Genomescan gene prediction, Similarity but not identity to protein.

FEATURES

source

Location/Qualifiers
 1..2281
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="CZECH II"
 /db_xref="taxon:10090"
 /clone="MGC:5966 IMAGE:5038182"
 /tissue_type="mammary tumor metastasized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMV."
 /clone_lib="NCI CGAP_Lu29"
 /lab_host="DH10B"
 /note="vector: pCMV-SPORT6"
 197..775
 /codon_start=1
 /product="Unknown (protein for MGC:5966)"
 /protein_id="AAH51053.1"
 /db_xref="GI:29835222"
 /translation="MQALKCVVGDGAVGKTCLLISYTNAPGEYPTVFDNYSANV
 MVDKPNLGLMDAGDEYDRLEPSYQDVELICFSLVSPASFNVRKAWPEYR
 HHCPTNLIIVGTDLDRDDTIKLEKRLITPTVQGLAMAKEIGAVKYLBCSL
 TORGLKTVFDEARVLCPPEVKRKKCLLL"
 BASE COUNT 568 a 569 c 504 g 640 t
 ORIGIN

Query Match 66.8%; Score 679.2; DB 10; Length 2281;

Best Local Similarity 86.1%; Pred. No. 6.7e-191;

Matches 876; Conservative 0; Mismatches 113; Indels 29; Gaps 10;

QY	1	CCCCATCTTCGTGCAGATTAAAGATTGCCAAATACCTTGTGAACCTAAGTTGGTGTG 60
Db	1019	CCCCATCTTCGTGCAGATTAAAGATTGCCAAATACCTTGTGAACCTAAGTTGTGTG 1078
QY	61	CTGAGAACAGCTAAGCTAGCTGTTGAGAGA--CTTTGCTCTTAAGAGACTGCAGC 117
Db	1079	CTGAGAACACTAAGCTAAGCTCTCTTGAGAGACTGCTGCTTAAGAGACTGCAGC 1138
QY	118	TTCTGGGCTCAGGGTGCAGACCCCTCCGCTAGCTCCAGACCCGCTGCACACAGCAGCC 177
Db	1139	TTCTGGAGTCAGGGGTGCAGACCCCTCCATAGTTCCAGACCCGCTGCACACAGCAGCC 1198
QY	178	TCCTTAATGACAGCTGCCATGTAAACGACCTGTAACTTATCAGCCCATGCTCATACGT 237
Db	1199	TCCTTAATGACAGCTGCCATGTAAACGACCTGTAACTTATCAGCCCATGCTCATACGT 1258
QY	238	AACCTTTGTACTGTACGTAC--GATGGGTGTAAACAGCTGCTGCTTTGATTTTATAGTG 294
Db	1259	AACCTTTGTACTGTACGTACAGTGATGAGTGTGCACCTCAGCTCTTGTTGATTCATAGT 1318
QY	295	AGTTCTCTAAATACAGCTGACCGGTTCTGAGCGGTTTGAACAGAACTCTGCCTCCTG 354
Db	1319	AGTTCTCTAAAGAACCCAGCGGCTAGCTTTTTCAGAGCTTTTGAACAGAACTCTGCTCCTG 1378
QY	355	TGTTGCTCTTAACGAGTATCTGCTCCTAGTGTGGGTGTGGGTGAGTGTGTGAA 414
Db	1379	TGTTGCTCTTAAGAGTATCTGCTCCTAGTGTGGGTGTGGGTGAGTGTGTGAA--GTGTGAA 1436
QY	415	ACACGAGCTCATCAAGGAGACAGACAGTATTTTGACTAATATGAGTACAGATTATTT 474
Db	1437	ACACGACATCATCAAGGAGACAGACAGTATTTTGACTAATATGAGTACAGATTATTT 1496
QY	475	ACACTACATTGTACATGAGTAAATTCACATGAATAAAGTGTACCGGTAAGCTTTTAA 534
Db	1497	ACACTACATTGTACATGAGTAAATTCACATGAATAAAGTGTACCGGTAAGCTTTTAA 1556

QY	535	ACGGTTAATTTCTGTCAAAACAGTAGATGACAAATAGCCGATCTTATCATAGTCTCTC----- 589
Db	1557	ACGGTTAATTTCTGTCAAAACAGTAGGTGACAAATAGCCAACTTATCATAGTCTCTTGTGAG 1616
QY	590	-----TCATTGAGCCCCCTTCCCCCTGCTGCTCCCTCCCGAGATGGGCGGTGAGTCCA 642
Db	1617	CATCCCTTCGGCGCCCCCACCCTCCACTGCGCTCCCTCCCGAGATGGGCGGTGAGTCCA 1676
QY	643	TATTTAAACTGGCCATCTCTCAAGTGTCTAACTTAGCAAGTGCCTTTCTTTAGGACCCCC 702
Db	1677	TATTTAAACTGGCCACCTCTACAGTTCTAACTTAGCAAGTGCCTTTCTTTA-GAACCCCC 1734
QY	703	TTCTTAACGAGCAATATGCTGACCTGTACTATAGATCTTTCTGATATGATTCG-CA 761
Db	1735	TTCTTAACGAGCAATATGCTGACCTGTACTATAGATCTTTCTGATATGATTCGCAAGGTTT 1794
QY	762	GATTTTGGTAGATAGTAGAAGTGCCTGCTCTTTTACCTTCTTCTTACTCAG-CTGA 820
Db	1795	TTTCTTTGGTAGCTCAGTAGAAGTGCCTGCTCTTTTCTGCTTACTTTTACTCAGAGCTA 1854
QY	821	CTAGTGTCTCCCTTTCGTTTCTAGTAACTGGGTGTAGAAATCAAGTGTCTGCGGTTTACA 880
Db	1855	GTTAGTGTCTTCTAGTTTCTAGCAACTAGTGTGCAATCATGTGTGTGAGCTTTAAG 1914
QY	881	GTTTTTAACTATTTTATGATTTCTGAAAC--ATCACTGTCTTCCAGAGTACCAACT 938
Db	1915	GTTTTTAACTATTTTATGATTTCTTAACTATGAACTTCTTAAACAGAGTACCAACT 1974
QY	939	GTCATGTGATTGANGCGCCCTCTAGACTTACCCACGGGACACATGCTTCGGT 996
Db	1975	GTCATGTGACTAATGCTG---CCTCTAGACCTCCCGACGAGACAGCTTCTCTGT 2029

RESULT 5

BC003828

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL	Submitted (28-FEB-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
REMARK COMMENT	NIH-MGC Project URL: http://mgc.nci.nih.gov Contact: MGC help desk Email: cgabs-r@mail.nih.gov Tissue Procurement: Gilbert Smith, Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305 Web site: http://www.shgc.stanford.edu Contact: (Dickson, Mark) mdcpaxil1.stanford.edu Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.		
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAX Plate: 11 Row: e Column: 22. Location/Qualifiers 1. .2319 /organism="Mus musculus" /mol_type="mRNA" /strain="FVB/N" /db_xref="taxon:10090" /clone="MGC:6235 IMAGE:3593957" /tissue="mammary tumor. Metallothionien-RGF alpha model. 10 month old virgin mouse. Taken by biopsy." /clone_lib="NCI CGAP_Maml" /lab_host="DH10E" /notes="Vector: pCMV-SPORT6"		
source	1. .2319 /organism="Mus musculus" /mol_type="mRNA" /strain="FVB/N" /db_xref="taxon:10090" /clone="MGC:6235 IMAGE:3593957" /tissue="mammary tumor. Metallothionien-RGF alpha model. 10 month old virgin mouse. Taken by biopsy." /clone_lib="NCI CGAP_Maml" /lab_host="DH10E" /notes="Vector: pCMV-SPORT6"		
gene	1. .2319 /gene="Rac1" /note="synonym: D5Brttd559e" /db_xref="LocusID:19353" /db_xref="MGI:97845" 198..776 /codon_start=1 /product="Rac1 protein" /protein_id="AAH03828.1" /db_xref="GI:13277918" /db_xref="LocusID:19353" /translation="MQAIKVVVGDAVGKICLLISYTNAPFGEVPTVFDNYSANV MVDKPNVLMDTAGEYDRLRPLUSYPTDVFLLICFLSPASENVRKAWPEVR HHCNPPIILVETKLDLDDKDTIEKLKKLPITYPOGLAWAKBIGAVKYLECSAL TQGLKTVFDEAIRVLCPPPVKKRRKCLLL"		
CDS	603 a 573 c 503 g 640 t		
BASE COUNT	603 a 573 c 503 g 640 t		
ORIGIN	Query Match 66.8%; Score 677.6; DB 10; Length 2319; Best Local Similarity 85.3%; Pred. No. 2e-190; Matches 870; Conservative 0; Mismatches 119; Indels 31; Gaps 9;		
QY	1	CCCTATTCTGCTCAGATTAAAGATTGCCAAATACCTTGTAACCTAAAGTTCGTTGTG 60	
Db	1020	CCCCCATTTCTGCTCAGATTAAATATGCCAAATACCTTCTGAACCTAAAGTTCGTTGTG 1079	
QY	61	CTGAGAACACGTAAGCACTAAGCTGTGTGAGAGA---CTTTGTCTTAAAGAGACTGCAGC 117	
Db	1080	CTGAGAACACCTAAGCACTAAACTCTCTTCAGAGACTTCTGTTGCTAAGAGACCGCAGC 1139	
QY	118	TTCTGGCTCAGGGTGCAGACCTCCCGTAGCTCCAGACCGGTGTGACACAGCAGACC 177	
Db	1140	TTCTGGAGCAGGGTGCAGACCTTCCAGAGTTCACAGACCGGTGTGACACAGCAGACC 1199	
QY	178	TCCTTAATGACAGCTGCCATGTAAACGACCTGTAACTTATCAGCCCATGCTCATTAAGT 237	
Db	1200	TCCTTAATGACAGCTGCCATGTAAACGACCTGTAACTTATCAGCCCATGCTCATTAAGT 1259	
QY	238	AACCTTTGACTAGCTGCAC---GATGGGTGTAACAGCTCTGCTCTTTGATTTCATAGTG 294	
Db	1260	AACCTTTGACTAGCTGCACAGTATGAGTGTGACAGCTCAGCTCTTTGATTTCATAGTG 1319	

QY	295	AGTTCTCTAAATACAGCTGACCGCTTCTCAGGCTTTGAACAGAACTCTGGTCCCTG 354	
Db	1320	AGTTTCTAAAGACACCGGACTAGCTTTTCAGACTTTGAACAGAACTCTGGTCCCTG 1379	
QY	355	TGTTGCTCTAAACGAAGTATTTCTGTTCTAGTCTGGGTGTGCTGGGTGAGTGTGAA 414	
Db	1380	TGTTGCCCTTAATGAAGTATTTCTGTTCTAGTGTGGGTGTGCTGGGTGGA--GTGTGAA 1437	
QY	415	ACACGACGTCATCAAGGAGACAGACAGTATTTTGACTTAATATGAAGTACAGATTATTT 474	
Db	1438	ACACGACATGATCAAGGAGACAGACAGTATTTTGACTTAATATGAAGTACAGATTACTTT 1497	
QY	475	ACACTACATTTGACATGGAGTAACTCACTGAATAAAGTGTCAACGGGTAAAGCTTTTAA 534	
Db	1498	ACACTACATTTGACATGGAGTAACTCACTGAATAAAGTGTCAACGGGTAAAGCTTTTAA 1557	
QY	535	ACGGTTAATTTCTGTCACACAGTAGATGACAAATGCCGATCTTATCAGTGTCTC----- 589	
Db	1558	ACGGTTAATTTCTGTCACACAGTAGATGACAAATGCCGATCTTATCAGTGTCTCCTTGAG 1617	
QY	590	-----TCTTGAGCCCCCTTCCCTCTGCTGCTCCCTCCAGATGGGGGTGAGTCCA 642	
Db	1618	CATCCCTTCGCGCCCCCACTCCACTGCGCTCCCTCCCTCCAGATGGGGGTGAGTCCA 1677	
QY	643	TATTTAACTGGCCATCTCTCAGAGTTGCTAACTTAGCAAGTGTCTTTCTTAGGACCCCC 702	
Db	1678	TATTTAACTGGCCATCTCTCAGAGTTGCTAACTTAGCAAGTGTCTTTCTTTA--GAACCCC 1735	
QY	703	TTCTTAACGAGCAATATGCTGACCTGTACTATAAGATCTTCTGTATAATGCA-----TT 757	
Db	1736	TTCTTAACGAGCAATATGCTGACCTGTACTATAAGATCTTCTGTATAACGACAGATTT 1795	
QY	758	CGGAGATTTTGTGTAGATAGTAGAAGTGGCTTCTCTGTTTTCACCTCTCTTACTCAG-- 816	
Db	1796	TTTGTTTTTTTTGTGTAGTCTAGTAGAAGTGGCTTCTCTGTTTTCACCTCTCTTACTCAGA 1855	
QY	817	CTGACTAGTGTCTCCCTTCGTTTCTAGTAACCTGGGTGTAGAAATACGCTGCTGGCGCTT 876	
Db	1856	GCTAGTGTGTCTCTCTAGTTTCTTAGTCTTAGTGTGCAATCATGTGTGACGCTT 1915	
QY	877	TACAGTTTTTAACTATTTTAGATATTTGAAACATCATCTGTCTTGCCAGAGTACCAACA 936	
Db	1916	TACGGTTTTTAACTATTTTAGATATTTTAACTATGAACCTTCTAAACAGTACCAACA 1975	
QY	937	CTGCTATGATGATGCGCCCTCTAGACCTCACCGCGGACACATGCTTCGCT 996	
Db	1976	CTGCTATGATGATGCGCTCTAGACCTCACCGCGGACACATGCTTCCTGT 2032	

AC068493 269081 bp DNA linear HTG 15-MAY-2002
Mus musculus clone RP23-76K1 strain C57BL6/J, WORKING DRAFT
SEQUENCE, 26 unordered pieces.
AC068493
AC068493.10 GI:15148081
HTG; HTGS PHASE1; HTGS DRAFT.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 269081)
Montgomery, K.T., Grills, G., Han, J., Lee, E., Long, J., Pomerantz, R., Ioshkhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A., Gordon, M., Goltz, J.S. and Kucherlapati, R.
High Throughput Mouse Sequencing
Unpublished
2 (bases 1 to 269081)
Montgomery, K.T., Grills, G., Han, J., Lee, E., Long, J., Pomerantz, R., Ioshkhes, I.P., Shim, C., Decker, J., Thomas, E., Perera, A., Gordon, M., Goltz, J.S. and Kucherlapati, R.
Direct Submission

RESULT 6
AC068493
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE

JOURNAL

Submitted (03-MAY-2000) Department of Molecular Genetics, Albert Einstein College of Medicine Genome Center, 1300 Morris Park Ave., Bronx, NY 10461, USA

COMMENT

On Aug 11, 2001 this sequence version replaced gi:14993654.

-----Genome Center

Center: Harvard Partners Genome Center

Center Code: HPGC

Web site: <http://www.hpcgg.org/Sequence/mouse.html>

Contact: hpgc@model.mgh.harvard.edu

-----Summary Statistics

Center project name: ABN

Sequencing vector: pUC18; L08752

Chemistry: Dye-terminator Big Dye; 100%

*Consensus quality: 256302 at least Q20

*Consensus quality: 254124 at least Q30

*Consensus quality: 250276 at least Q40

Estimated insert size: agarose-FP - N/A

**Estimated insert size: 268581 - sum-of-contigs

Quality coverage: agarose-FP - N/A

Quality coverage: 6.2 x in Q20 bases; sum-of-contigs estimation

* NOTE: This is a 'working draft' sequence. It currently consists of 26 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 51730: contig of 51730 bp in length
* 51731 51750: gap of unknown length
* 51751 82423: contig of 30673 bp in length
* 82424 82444: gap of unknown length
* 82444 118158: contig of 35715 bp in length
* 118159 118179: gap of unknown length
* 118179 139441: contig of 21263 bp in length
* 139442 139461: gap of unknown length
* 139462 160638: contig of 21177 bp in length
* 160639 160659: gap of unknown length
* 160659 177612: contig of 16953 bp in length
* 177612 177632: gap of unknown length
* 177632 190991: contig of 13360 bp in length
* 190992 191011: gap of unknown length
* 191012 200128: contig of 9117 bp in length
* 200129 200149: gap of unknown length
* 200149 211882: contig of 11734 bp in length
* 211883 220773: contig of 8871 bp in length
* 220774 220794: gap of unknown length
* 220794 229643: contig of 8850 bp in length
* 229644 229664: gap of unknown length
* 229664 238339: contig of 8676 bp in length
* 238340 238359: gap of unknown length
* 238360 244848: contig of 6489 bp in length
* 244849 244869: gap of unknown length
* 244870 249725: contig of 4857 bp in length
* 249726 249746: gap of unknown length
* 249746 252523: contig of 2778 bp in length
* 252524 252544: gap of unknown length
* 252544 257520: contig of 4977 bp in length
* 257521 257541: gap of unknown length
* 257541 260396: contig of 2856 bp in length
* 260397 260417: gap of unknown length
* 260417 262709: contig of 2293 bp in length
* 262710 262730: gap of unknown length
* 262730 263954: contig of 1205 bp in length
* 263955 264196: gap of unknown length
* 264197 264217: contig of 242 bp in length
* 264218 264629: gap of unknown length
* 264630 264649: contig of 413 bp in length
* 264650 265608: contig of 959 bp in length
* 265609 265628: gap of unknown length

```

```

* 265629 265935: contig of 1307 bp in length
* 265936 265955: gap of unknown length
* 265956 268335: contig of 1380 bp in length
* 268336 268355: gap of unknown length
* 268356 268549: contig of 194 bp in length
* 268550 268569: gap of unknown length
* 268570 269081: contig of 512 bp in length.

FEATURES             Location/Qualifiers
     source            1..269081
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL6/J"
                     /db_xref="taxon:10090"
                     /clone="RP23-76K1"
                     /sex="male"
     misc_feature       1..51730
                     /note="assembly_name:Contig138"
     misc_feature       51751..82423
                     /note="assembly_name:Contig137"
     misc_feature       82444..118158
                     /note="assembly_name:Contig136"
     misc_feature       118179..139441
                     /note="assembly_name:Contig135"
     misc_feature       139462..160638
                     /note="assembly_name:Contig134"
     misc_feature       160659..177611
                     /note="assembly_name:Contig133"
     misc_feature       177632..190991
                     /note="assembly_name:Contig132"
     misc_feature       191012..200128
                     /note="assembly_name:Contig131"
     misc_feature       200149..211882
                     /note="assembly_name:Contig130"
     misc_feature       211903..220773
                     /note="assembly_name:Contig129"
     misc_feature       220794..229643
                     /note="assembly_name:Contig128"
     misc_feature       229664..238339
                     /note="assembly_name:Contig127"
     misc_feature       238360..244848
                     /note="assembly_name:Contig126"
     misc_feature       244869..249725
                     /note="assembly_name:Contig125"
     misc_feature       249746..252523
                     /note="assembly_name:Contig124"
                     clone_end:SP6
                     vector_side:left
     misc_feature       252544..257520
                     /note="assembly_name:Contig123"
     misc_feature       257541..260396
                     /note="assembly_name:Contig122"
     misc_feature       260417..262709
                     /note="assembly_name:Contig121"
     misc_feature       262730..263954
                     /note="assembly_name:Contig120"
     misc_feature       263955..264196
                     /note="assembly_name:Contig119"
     misc_feature       264217..264629
                     /note="assembly_name:Contig118"
     misc_feature       264650..265608
                     /note="assembly_name:Contig117"
     misc_feature       265629..266935
                     /note="assembly_name:Contig116"
     misc_feature       266956..268335
                     /note="assembly_name:Contig115"
     misc_feature       268356..268549
                     /note="assembly_name:Contig114"
     misc_feature       268570..269081
                     /note="assembly_name:Contig113"
BASE COUNT    69726 a 66027 c 64672 g 67724 t    932 others
ORIGIN

```

Query Match

66.8%; Score 677.6; DB 2; Length 269081;

Best Local Similarity 85.3%; Pred. No. 3e-190;			
Matches 870; Conservative 0; Mismatches 119; Indels 31; Gaps 9;			
QY	1	CCCCATTCTTCTCAGATTAGAAATGCGAAATACCTTGTGAACATAAGTTGGCTGTG	60
Db	62207	CCCCATTCTTCTCAGATTAGAAATGCGAAATACCTTGTGAACATAAGTTGGCTGTG	62266
QY	61	CTGAGAACGTAAGCACTAAGCTGTGAGAG---CTTTGTCCTTTAAGAACTGCAGC	117
Db	62267	CTGAGAACCACTAAGCACTAAGCTGTGAGAG---CTTTGTCCTTTAAGAACTGCAGC	62326
QY	118	TTCTGGGCTCAGGGTGACAGCCCTCCCTAGTCTCCAGACCGTGTGACACAGCAGCC	177
Db	62327	TTCTGGGCTCAGGGTGACAGCCCTCCAGAGTTCAGAGTTCAGACCGTGTGACACAGCAGCC	62386
QY	178	TCCTTAATGACAGCTGCCATGTAACGCACTGTAACTTATCAGCCCATGCTCATTAAGT	237
Db	62387	TCCTTAATGACAGCTGCCATGTAACGCACTGTAACTTATCAGCCCATGCTCATTAAGT	62446
QY	238	AACTTTGTTACTTACGTAC---GATGGGTGAACAGCTCTCTCTTGTATTTATAGTG	294
Db	62447	AACTTTGTTACTTACGTACGTGATGAGTGTGACAGCTCAGCTCTTGTATTTATAGTG	62506
QY	295	AGTTCTCTAAATATACAGCTGACCGGCTTCTGACGCTTTTGAACAGAACTCTGGCTCCCTG	354
Db	62507	AGTTCTCTAAATATACAGCTGACCGGCTTCTGACGCTTTTGAACAGAACTCTGGCTCCCTG	62566
QY	355	TGTTGCTCTAAAGAAATTTCTGTTCTTGTAGTGGGTGGTGGTGGTGGTGGTGGTAA	414
Db	62567	TGTTGCTCTAAATGAAATTTCTGTTCTTGTAGTGGGTGGTGGTGGTGGTGGTGGTAA	62624
QY	415	ACACGAGCTCATCAAGGACACAGACAGTATTTTGTACTTAATATCAAGTAGAGTTAATTT	474
Db	62625	ACACGAGCTCATCAAGGACACAGACAGTATTTTGTACTTAATATCAAGTAGAGTTAATTT	62684
QY	475	ACACTACATGTTACATGGAGTAACTCAACTGAATAAAGTGTACCGGTAAAGCTTTTAA	534
Db	62685	ACACTACATGTTACATGGAGTAACTCAACTGAATAAAGTGTACCGGTAAAGCTTTTAA	62744
QY	535	AGGTTAATTTCTGTCACAGTAGATGACAAATGGCGGATCTTATCAGTGTCTC-----	589
Db	62745	AGGTTAATTTCTGTCACAGTAGATGACAAATGGCGGATCTTATCAGTGTCTTCTAG	62804
QY	590	-----TCTTGAGCCCCCTCCCTCCCTGCTGCTCCCTCCAGATGGGGTGTGAGTCCA	642
Db	62805	CATCCCTTCGGCCCCCACCCTCCACTGCGGCTCCCTCCAGATGGGGTGTGAGTCCA	62864
QY	643	TATTTAACTGGCCATCCTCACAGTTGCTAACTTAGCAAGTGTCTTTTCTTTAGGACCCCT	702
Db	62865	TATTTAACTGGCCATCCTCACAGTT-CCTAACTTAGCAAGTGTCTTTTCTTTA-GAACCCCT	62922
QY	703	TTCTTAACGAGCAATATGCTGACCTGTACTATAAGATCTTCTGATAATGCA-----TT	757
Db	62923	TTCTTAACGAGCAATATGCTGACCTGTACTATAAGATCTTCTGATAACGCAAGATTT	62982
QY	758	CGGAGATTTTGTGGTAGATAGTAGAGTGGTTCCTGTTTCCACCTCTCTTTACTCAG-	816
Db	62983	TTTGTGTTTTTGTGGTAGCTCAGTAGAGTGGTTCCTGTTTCCGCTTACTTACTCAGA	63042
QY	817	CTGACTAGTGTCTTCCCTCTGTTTCTAGTAACCTGGGTGTAGAAATCAGCTGTGCGGCTT	876
Db	63043	GCTAGTGTGCTTCTTCTAGTCTTCTAGCAACTAGGTGTGCAATCATGTGTTGCGCTT	63102
QY	877	TACAGTTTTTAACTATTTAGATATTTCTGAACATCATCTGTCTTCCGAGTACCAACA	936
Db	63103	TACGGTTTTTAACTATTTAGATATTTCTTAAACTATGAACTTCTTAAACAGTACCAACA	63162
QY	937	CTGTGATGCTAATGATGCCGCCCTCTAGCACTCACCACGCGGACACATGCTTCCGTT	996
Db	63163	CTGTGATGCTAATGCTG---CCTCTAGACTTCCGCGGACGAGCAGCTTCTCTGT	63219

RESULT 7

AC102775
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AC102775 185245 bp DNA linear HTG 23-MAR-2003
Mus musculus clone RP23-115C10, WORKING DRAFT SEQUENCE, 9 unordered pieces.
AC102775
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Birren, B., Nusbaum, C. and Lander, E.
1 (bases 1 to 185245)
Mus musculus, clone RP23-115C10
Unpublished
2 (bases 1 to 185245)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faru, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 185245)
Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, S., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cooke, A., Cooke, P., Corum, B., DeArellano, K., Faru, S., Diaz, J. S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faru, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, P., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (23-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 23, 2003 this sequence version replaced gi:22381790.
All repeats were identified using RepeatMasker:
Smit, A. F. A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L19647
 Center clone name: 115_C.10
 ----- Summary Statistics
 Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 183758 bases at least Q40
 Consensus quality: 184136 bases at least Q30
 Consensus quality: 184291 bases at least Q20
 Insert size: 178000; agarose-fp
 Insert size: 184445; sum-of-contigs
 Quality coverage: 12.6 in Q20 bases; agarose-fp
 Quality coverage: 12.2 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 9 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 53693: contig of 53693 bp in length
 * 53694 53793: gap of 100 bp
 * 53794 56675: contig of 2882 bp in length
 * 56675 56775: gap of 100 bp
 * 56775 60589: contig of 3814 bp in length
 * 60589 60689: gap of 100 bp
 * 60689 67210: contig of 6521 bp in length
 * 67210 67310: gap of 100 bp
 * 67310 82585: contig of 15275 bp in length
 * 82585 82685: gap of 100 bp
 * 82685 103118: contig of 20433 bp in length
 * 103118 103218: gap of 100 bp
 * 103218 132145: contig of 28927 bp in length
 * 132145 132245: gap of 100 bp
 * 132245 182335: contig of 50090 bp in length
 * 182335 182435: gap of 100 bp
 * 182435 185245: contig of 2810 bp in length.
 * 182436 185245: contig of 2810 bp in length.
 * Location/Qualifiers
 * 1..185245
 * /organism="Mus musculus"
 * /mol_type="genomic DNA"
 * /db_xref="taxon:10090"
 * /clone="RP23-115C10"
 * /clone_lib="RPCI-23 Female Mouse BAC"
 * 1..53693
 * /note="assembly_fragment
 * clone_end:sp6
 * vector_side:left"
 * 53794..56675
 * /note="assembly_fragment"
 * 56775..60589
 * /note="assembly_fragment"
 * 60689..67210
 * /note="assembly_fragment"
 * 67311..82585
 * /note="assembly_fragment"
 * 82686..103118
 * /note="assembly_fragment"
 * 103219..132145
 * /note="assembly_fragment"
 * 132246..182335
 * /note="assembly_fragment"
 * 182436..185245
 * /note="assembly_fragment
 * clone_end:r7
 * vector_side:right"
 *
 * BASE COUNT 56870 a 34956 c 35069 g 57550 t 800 others
 * ORIGIN

Query Match 39.6%; Score 403.2; DB 2; Length 185245;
 Best Local Similarity 89.7%; Pred. No. 1.9e-108;
 Matches 481; Conservative 0; Mismatches 43; Indels 12; Gaps 4;
 QY 1 CCCCTATTCTTGTCTAGATTAAAGAAATGCGCAAAATACCTTGTGAACCTAAAGTTGGCTTGTG 60
 DB 74 CCCCTATTCTTGTCTAGATTAAAGAAATGCGCAAAATACCTTGTGAACCTAAAGTTGGCTTGTG 133
 QY 61 CTGAGAACACGTAAGCACTAAGCTGTGTTGAGAGA---CTTTGTCTCTTAAGAAAGATGCGAGC 117
 DB 134 CTGAGAACACCTTAAGCACTAAGCTGTGTTGAGAGACTTCTGTTGCTTAAGAAAGATGCGAGC 193
 QY 118 TTCTGGGCTCAGGGGTGAGACCCCTCCGGTAGCTCCAGAGCCGCTGTGACACAGACACGCC 177
 DB 194 TTCTGGGAGTCAGGGGTGAGACCCCTCCATAGTTCCTCCAGACAGTGTGACACAGACACGCC 253
 QY 178 TCCTTAATGACACGCTGCCATGTAAACCACTGTAACTTATCAGCCCATGCTCAATTACGT 237
 DB 254 TCCTTAATGACACGCTTGCATGTAAACCACTGTAACTTATCAGCCCATGCTCAATTACGT 313
 QY 238 AACTTTCTACTGTACGTCAAC---GATGGGTGTAAACAGCTCTGCTCTTTGATTTCATAGT 294
 DB 314 AACTTTTACTGTACGTCAACGTGATGAGTGTGACAGCTCAGCTCTTTGATTTCATAGT 373
 QY 295 AGTTCTCTAAATACCACTGACCGGCTTCTGAGGCTTTGAAACAGAACTCTGCTCCTG 354
 DB 374 AGTTTCTTAAAGACCACTGACCTAGCTTTGACAGACTTTGAAACAGAACTCTGCTCCTG 433
 QY 355 TGTTCCTCTTAAACGAAGTATTCTGCTCTAGTCTGCTGGTGTGCTGGGTGAGTGTGTA 414
 DB 434 TGTTCCTCTCTAAGTATTCTGCTCTAGTGTGCTGGTGTGCTGGGTGGA--GTGTGAA 491
 QY 415 ACAGCAGCTCATCAAAAGG---AGACAGACAGATATTTTGACTAATATGAAGTAGAGATTA 470
 DB 492 ACAGCAGCATCAAAAGGAGACAGACAGACAGATATTTTGACTAATATGAAGTAGAGATTA 551
 QY 471 ATTACACTACATGTACATGGAGTATTTCAACTGAATAAAGTGTCAACGGGTAAA 526
 DB 552 CTTTACACTACATTTGTACATGGAGTATTTCAACTGAATAAAGTGTCAATGGGGAA 607
 RESULT 8
 AL583884 202342 bp DNA linear ROD 24-DEC-2002
 LOCUS Mouse DNA sequence from clone RP23-324B16 on chromosome 15,
 DEFINITION complete sequence.
 ACCESSION AL583884
 VERSION AL583884.20 GI:27368255
 KEYWORDS HTG.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 202342)
 Smith,M.
 Direct Submission
 Submitted (21-DEC-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Dec 23, 2002 this sequence version replaced gi:26788018.
 Sequence from the Mouse Genome Sequencing Consortium whole genome
 shotgun may have been used to confirm this sequence. Sequence data
 from the whole genome shotgun alone has only been used where it has
 a phred quality of at least 30.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: humquery@sanger.ac.uk

 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em.: EMBL; Sw.: SWISSPROT; Tr.: TRMBL; Wp.: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-324B16 is from the RPCI-23 Mouse PAC Library constructed by the group of Pieter de Jong.

For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6.

FEATURES

source

Location/Qualifiers

1..202342
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="15"
/clone="RP23-324B16"
/clone_lib="RPCI-23"

BASE COUNT 60264 a 39265 c 39460 g 63353 t

ORIGIN

Query Match 39.6%; Score 403.2; DB 10; Length 202342;

Best Local Similarity 89.7%; Pred. No. 1.9e-108; Indels 12; Gaps 4;

Matches 481; Conservative 0; Mismatches 43;

QY 1 CCCTATTCTGCTCAGATTAGAAATGCGAAATACCTTGTGAATTAAGTTCGTTGTG 60

Db 184518 CCCTATTCTGCTCAGATTAGAAATGCGAAATACCTTGTGAATTAAGTTCGTTGTG 184577

QY 61 CTGAGAACGTTAGACCTAAGCTGTGTGAGAGA---CTTGTCTTAAAGAACTGCAGC 117

Db 184578 CTGAGAACACCTAAGACCTAAGCTGTGTGAGAGA---CTTGTCTTAAAGAACTGCAGC 184637

QY 118 TTCTGGGCTCAGGGTGACAGCCCTCCGCTAGCTCCAGACCGTGTGACACACAGCC 177

Db 184638 TTCTGGAGTCAGGGTGACAGCCCTCCGCTAGCTCCAGACCGTGTGACACACAGCC 184697

QY 178 TCCTTAATGACCGCTGCCATGTACCGACCTCTAATCTATCAGCCCATGCTCATTACGT 237

Db 184698 TCCTTAATGACCGCTGCCATGTACCGACCTCTAATCTATCAGCCCATGCTCATTACGT 184757

QY 238 AACTTTGTACTGTACGTAC---GATGGGTGTAAACAGCTCTGCTTTGATTTTCATAGT 294

Db 184758 AACTTTTATCTGTACGTACAGTGTAGTGTGACAGCTCAGCTCTTTGATTTTCATAGT 184817

QY 295 AGTTCTCTAAATACACAGCTGACCGGCTTCTGACGGCTTTGAACAGAACTCTGGCTCCTG 354

Db 184818 AGTTTCTAAAGACACCGCACTAGCTTTTGACAGCTTTGACAGAACTCTGGTTCGTG 184877

QY 355 TGTGTGCTCTAACGAAGTATCTGTTCCTAGTCTGGGTGTGCTGGGTGAGTGTGTGAA 414

Db 184878 TGTGTGCTCTCATGAAGTATCTGTTCCTAGTCTGGGTGTGCTGGGTGGA--GTGTGAA 184935

QY 415 ACACGAGCTCATCAAGG---AGACAGACAGTATTTTGAATAATTAAGTACAGATTA 470

Db 184936 ACACGACCATCAAGGAGACAGACAGATTTTGAATAATTAAGTACAGATTA 184995

QY 471 ATTATACACTACATTTGACATGAGTAATTCAACTGAATAAAAGTGCACGGGTAAA 526

Db 184996 CTTTACACTACATTTGACATGAGTAATTCAACTGAATAAAAGTGCACGGGTAAA 185051

RESULT 9

AC126556

LOCUS

DEFINITION

AC126556

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

AC126556 179685 bp DNA linear ROD 11-JUN-2003
Mus musculus chromosome 15 clone RP23-64L3, complete sequence.

AC126556 AC126556.4 GI:31581824
HTG.

Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

Wilson,R.K.
The sequence of Mus musculus clone

Unpublished
2 (bases 1 to 179685)

McPherson,J.D. and Waterston,R.H.
Direct Submission

Submitted (06-JUL-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

3 (bases 1 to 179685)

McPherson,J.D. and Waterston,R.H.
Direct Submission

Submitted (07-JAN-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

4 (bases 1 to 179685)

Wilson,R.K.
Direct Submission

Submitted (11-JUN-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

On Jun 11, 2003 this sequence version replaced gi:27502524.

----- Genome Center -----

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu>

Contact: submissions@watson.wustl.edu

----- Project Information -----

Center project name: M BA0064L03

----- Location/Qualifiers -----

1..179685

/organism="Mus musculus"

/mol_type="genomic DNA"

/db_xref="taxon:10090"

/chromosome="15"

/clone="RP23-64L3"

BASE COUNT 47439 a 39048 c 39639 g 53560 t

ORIGIN

Query Match 33.2%; Score 337.2; DB 10; Length 179685;

Best Local Similarity 84.5%; Pred. No. 8.9e-89;

Matches 443; Conservative 0; Mismatches 63; Indels 18; Gaps 5;

QY 16 AGATTAAAGTAATGCCAAAATA-CTTGTGAATAAGTTGCTGTGCTGAGACACAGTAA 74

Db 56268 AAAAAAAGTAATGCCAAAATAACCTTCTGAACCTGAAGTTGCTGTGCTGAGAACACCTAA 56327

QY 75 GCATTAAGTGT-----TGAGAGACTTTGCTTAAAGAACTGCAGCTTCCTGG-----GC 125

Db 56328 GCATTAAGTGTCTTTGAGAGACTTCTGCTGCTGAAGAGGACTGCAGCTTCTGAGCCAGG 56387

QY 126 TCAGGGGTGCAGACCCCTCCCGTAGCTCCAGAGCCGCTGTGACAGACACAGCAGCCCTTAAT 185

Db 56388 TCAGGGGTGCAGCCCTTCCAGAGTCCCGAGCCGCTGTGACAGACACAGCAGCCCTTAAT 56447

QY 186 GACAGCTGCCATGTAAACGACCTGTAACTTATCAGCCCATGCTCATACGTAACTTTGT 245

Db 56448 GACAGCTGCCATGTAAACGACCTGTAACTTATCAGCCCATGCTCATACGTAACTTTGT 56507

QY 246 ACTGTACGTCAC---GATGGGTGTAAACAGCTGCTCTTTGATTTCATAGTGTGTTCTCT 302

Db 56508 ACTGTACGTCACAGTGTGAGCTGAGCTCTTCGATTTTCAATGAGTGTCTCT 56567

```
QY 303 AAAATACCAGCTGACCGGCTTCTGAGGCTTTTGAACAGAACTCTGCGCTCCCTGTTGGCT 362
Df 56568 AAAAGACCGGCGATAGCTTTTGCAGAGCTTTGAACAGATCTCTGTTCCCTGTTGGCT 56627
QY 363 CTAACGAAGTATTCCTAGTCTGCTGGTGTGCTGGTGGAGTGTGTGAACACGACG 422
Df 56628 CTAATGAAGTATTCCTAGTCTGCTGGTGTGCTGGTGGAGTGTGAACACGACATG 56687
QY 423 TCATCAAGGAGACAGACAGTATTTTGAAGTATTAAGTATTAAGTATTAAGTATTAAGT 482
Df 56688 ATCAAGGAGACAGACAGTATTTTGAAGTATTAAGTATTAAGTATTAAGTATTAAGT 56742
QY 483 TTGTACATGGAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGT 526
Df 56743 TTGTACATGGAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGT 56786

RESULT 10
AX163752 1022 bp DNA linear PAT 22-JUN-2001
LOCUS
DEFINITION Sequence 16 from Patent WO0138579.
ACCESSION AX163752
VERSION AX163752.1 GI:14544858
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Gould-Rothberg,B.E., Dipippo,V.A., Ramseh,T.M. and Gerwein,R.W.
TITLE Method of identifying toxic agents using nsaid-induced differential
JOURNAL Gene expression in liver
Patent: WO 0138579-A 16 31-MAY-2001;
Curagen Corporation (US)
FEATURES
source
1. 1022
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 273 a 208 c 205 g 336 t
ORIGIN

Query Match 28.6%; Score 290.6; DB 6; Length 1022;
Best Local Similarity 64.5%; Pred. No. 4.5e-75;
Matches 664, Conservative 0; Mismatches 304; Indels 61; Gaps 13;

QY 1 CCCCATTCTGTTGATTAAGAGTTGCCAAATATACCTTCTGAACCTACACTGATTTGTT 57
Df 1 CCCCATTCTGTTGATTAAGAGTTGCCAAATATACCTTCTGAACCTACACTGATTTGTT 60
QY 58 GTGCTGAGAACACGCTAAGCCTAAGCTGTTGAGAGACTTTGCTCTTAAGAGACTGAGC 117
Df 61 GTGCCGAGAACACGCTAAGCCTAAGCTGTTGAGAGACTTTGCTCTTAAGAGACTGAGC 120
QY 118 TTCTGGGCTCAGG-GGTGACAGCCCTCCCGTAGC-----TCCAGACCGTGTGACACA 169
Df 121 TTCTGAGTTAGAGGTGACAGACTTGCTCTCTATGATTTCTAGATGCGTAAGACA 180
QY 170 GCACAGCCTCTTAATGACAGCTGCCATGTAAAGCAGCCTGTAACTTATCAGCCCATGCT 229
Df 181 GAACAGCCTCCGGAATGAAGCTTGGCCATGCACTCAACAGTGTAGTGTAGCAGCAGTGT 240
QY 230 CATTACCTAAGTGTGATGCTACGCTACGATGGGTGTAAACAGCTCTGCTCTTGAATTC 289
Df 241 CCGACATACATTTGATCTGATTAAGTGTAGGAGTGTAGCAGCTCAGCTCTTGGATCAGTC 300
QY 290 TAGTGAGTTCTCTAAATACAGCTGACCGCTTCTGAGGCTTTGAACAGACTCTGGC 349
Df 301 TTGTGATTTCTATGACGAGTTTCTGACAGCTTTTGGGAGATTTTGAACAGACTG-- 358
QY 350 TCCTGTGTTGCTCTTAACAGGATTTCTGTTCTTCTAGTCTGGGTGTGCTGGGTGAGTGT 409
Df 359 ----CTATTCTCTATGAGATTTCTGTT--TAGCTGTGGGTGTGCTGGGTGAGTGT 412
```

```
QY 410 GTGAACACGAGCTCATCAAGGAGACAGACAGTATTTTGAC-TAATATGAAGTAGAGAT 468
Df 413 GT-----GTGATCAAAAGGACAAAGACAGTATTTTGACAAATATCAAGTAGAG-- 460
QY 469 TAAATTACACTACATTTGACATGAGTAA--TTCACTGATTAAGTGTGACGGGTAA 525
Df 461 --ATTACACTACATTTGACATGAGTAAAGTGTGACGGGTAAAGTCTTAAAGGTGA 518
QY 526 AGCTTTTAAACGGTAAATTTCTGTCAACACAGTAGATGACAAATGGCCGATCTTATCAGTG 585
Df 519 ATTCTGTCAATGACAGTAGATGATGAAGAAAGGTTGGTATTATCAGAAATGTTTCT 578
QY 586 TCTCTTTGAGCCCCCTTCCCCCTGCTGCTCCCTCCAGATGGGCGTGTGAGTCCATAT 645
Df 579 TAAGCTTTTCTCTTACACCTGCCATGCTCCCAATTTGGGCATTTAATTCATCT 638
QY 646 TTAACCTGGCCATCTCAGCTTCTAAGTGTGAGTGTGCTTTCTTTAGGACCCCTTC 705
Df 639 TTAACCTGGTGTGCTGTTAGTCTGCTAAGTGTGCTTTCTTTATAGAACCCCTTC 698
QY 706 TTAACGAGCAATATGCTGACCTGTACTATAAGATCTTTCTGATAATGATTCGGAGATT 765
Df 699 TGACTGAGCAATATGCTTCTGATTAATAATCTTTCTGATAATGATTCGGAGATT 757
QY 766 TTTTGGTAGATGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 825
Df 758 TTTTGGTAGATGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 812
QY 826 GCTTCCCTCTGCTTTCTAGTAACTGGGTGAGTAAATCAGCTGCTGCTGCTGCTGCTGCTGCT 885
Df 813 GCTTCCCTCTGCTTTCTAGTAACTGGGTGAGTAAATCAGCTGCTGCTGCTGCTGCTGCTGCT 872
QY 886 TAAACTATTTTAGATA-----TTCTGAAACATCATCTGCTGCTGCTGCTGCTGCTGCTGCT 928
Df 873 TAAACTATTTTAGATAATTTCTAATGATGACCTCTTAACTACATCTGCTGCTGCTGCTGCT 932
QY 929 TACCAACACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 988
Df 933 TACGACACTGTCACTTGAACCACTG-ACCCTCTTTACCTGCTGCTGCTGCTGCTGCTGCTGCT 991
QY 989 CTTCGGGTA 997
Df 992 CCTCCTGTA 1000
```

```
RESULT 11
BC050687 2302 bp mRNA linear PRI 11-APR-2003
LOCUS
DEFINITION Homo sapiens, ras-related C3 botulinum toxin substrate 1 (rho
family, small GTP binding protein Rac1), clone MGC:60264
IMAGE:6149377, mRNA, complete cds.
ACCESSION BC050687
VERSION BC050687.1 GI:29792301
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2302)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (08-APR-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC/DCTP/DTF
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
```

Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www.shgc.stanford.edu>
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK plate: 110 Row: k Column: 14
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 9845510.

FEATURES
source

Location/Qualifiers
1..2302
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:60264 IMAGE:6149377"
/tissue type="Skin, melanotic melanoma."
/clone_lib="NIH MGC_72"
/lab_host="DH10B"
/note="vector: pCMV-SPORT6"
189..767
/codon_start=1
/product="ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1)"
/protein_id="AAH50887.1"
/db_xref="GI:29792302"
/db_xref="LocusID:5879"
/translation="MQAICQVVGDAVGKTKLLISYTNAPFGEYIPVFDNYSANV
MVDKPNVLGLMDTAQEDYDPLSPQTDVFLICFSLVSPSPENVRKAWYPPVR
HICNPNTILVGLKDRDDKPTIEKLEKLLPTIYVQGLMAKBIKAVKYLECSAL
TQGLKTVFDEAIRLVLCPPVKKRKKLLL"
BASE COUNT 617 a 524 c 498 g 663 t

CDS

Query Match 28.6%; Score 290.6; DB 9; Length 2302;
Best Local Similarity 64.5%; Pred. No. 4.8e-75;
Matches 664; Conservative 0; Mismatches 304; Indels 61; Gaps 13;
QY 1 CCCTATTCTGTCAGATTAGAAATGCGCAAAATACCTTGTGAAGTGC---GTT 57
Db 1050 CCCCATTCTGTTTCAGATTAGAGTTTGCAGAAATACCTTGTGAAGTGCATTGTT 1109
QY 58 GTGCTGAGACAGTAAAGCACTAGCTTGTGAGAGCTTTGCTTAAAGAGCTGAGC 117
Db 1110 GTGCGAGAACACCGGACCTGAACTTGTCAAGAGACCTTGTCTTTGAGAAGACGGTAGC 1169
QY 118 TTCTGGGCTCAGG-GGTGACAGACCTCCCGTAGC-----TCCGACCGGTGTGACACA 169
Db 1170 TTCTGCACTTAGGAGGTGACAGACTGTCTCTCTATGTAGTTCTCAGATGCGTAAAGCA 1229
QY 170 GCACAGCTCTTAAATGACAGCTGCGATGATGACGACCTGTAACTTATCAGCCCATGCT 229
Db 1230 GAACAGCTCCCGAATGAGCGTTGCAATGAACTACCACTGAGTTAGCAGACAGCTGTT 1289
QY 230 CATTAGCTAACTTGTACTGCTACGATGAGTGGGTGTAACAGCTCTCTCTTTGATTCA 289
Db 1290 CCGACATAACATGTACTGTAAATGGAGTGGCGGTAGCAGCTCAGCTCTTTGGATCAGTC 1349
QY 290 TAGTGAAGTTCTCTAAATAATACAGCTGACCGGCTTCTGAGCGTTTGAACAGAACTCTGGC 349
Db 1350 TTTGTGATTTATGACGAGTTTCTGACAGCTTTTGGGAGATTTTGAACAGAACTG-- 1407
QY 350 TCCTGTGCTGCTCTAAGAGATTTCTGTTCTAGTCTGGGTGCTGCTGGTGGATGT 409
Db 1408 -----CTATTTCTCTAATGAAGATTTCTGTT--TAGCTGTGGGTGTGCCGGGTGGGTGT 1461
QY 410 GTGAACACAGCGTCATCAAGGAGAGACAGAGTATTTTGAC-TAAATATGAAGTAGAGAT 468
Db 1462 GT-----GTGATCAAGGACAAGACAGATTTTTCACAAAATACGAGTGGAG-- 1509
QY 469 TAATTTACACTACATTTGATGAGTAA---TTCAACTGAATAAAGTGTACCGGTAA 525

BASE COUNT
ORIGIN

QY 1 CCCTATTCTGTCAGATTAGAAATGCGCAAAATACCTTGTGAAGTGC---GTT 57
Db 1050 CCCCATTCTGTTTCAGATTAGAGTTTGCAGAAATACCTTGTGAAGTGCATTGTT 1109
QY 58 GTGCTGAGACAGTAAAGCACTAGCTTGTGAGAGCTTTGCTTAAAGAGCTGAGC 117
Db 1110 GTGCGAGAACACCGGACCTGAACTTGTCAAGAGACCTTGTCTTTGAGAAGACGGTAGC 1169
QY 118 TTCTGGGCTCAGG-GGTGACAGACCTCCCGTAGC-----TCCGACCGGTGTGACACA 169
Db 1170 TTCTGCACTTAGGAGGTGACAGACTGTCTCTCTATGTAGTTCTCAGATGCGTAAAGCA 1229
QY 170 GCACAGCTCTTAAATGACAGCTGCGATGATGACGACCTGTAACTTATCAGCCCATGCT 229
Db 1230 GAACAGCTCCCGAATGAGCGTTGCAATGAACTACCACTGAGTTAGCAGACAGCTGTT 1289
QY 230 CATTAGCTAACTTGTACTGCTACGATGAGTGGGTGTAACAGCTCTCTCTTTGATTCA 289
Db 1290 CCGACATAACATGTACTGTAAATGGAGTGGCGGTAGCAGCTCAGCTCTTTGGATCAGTC 1349
QY 290 TAGTGAAGTTCTCTAAATAATACAGCTGACCGGCTTCTGAGCGTTTGAACAGAACTCTGGC 349
Db 1350 TTTGTGATTTATGACGAGTTTCTGACAGCTTTTGGGAGATTTTGAACAGAACTG-- 1407
QY 350 TCCTGTGCTGCTCTAAGAGATTTCTGTTCTAGTCTGGGTGCTGCTGGTGGATGT 409
Db 1408 -----CTATTTCTCTAATGAAGATTTCTGTT--TAGCTGTGGGTGTGCCGGGTGGGTGT 1461
QY 410 GTGAACACAGCGTCATCAAGGAGAGACAGAGTATTTTGAC-TAAATATGAAGTAGAGAT 468
Db 1462 GT-----GTGATCAAGGACAAGACAGATTTTTCACAAAATACGAGTGGAG-- 1509
QY 469 TAATTTACACTACATTTGATGAGTAA---TTCAACTGAATAAAGTGTACCGGTAA 525

Db 1510 --ATTACACTACATTGTACAAAGGAATGAAGTGTACCGGTAAAAAAGTTA 1567
QY 526 AGCTTTTAAACGGTTAACTTTCTGTCAAACAGTAGATGACAAATGCGCGATCTTATCAGTG 585
Db 1568 ATTTCTGTCAAATGACAGTAGATGAGAAAGAAAGTTGGTATTATCAGAAATGTTTCT 1627
QY 586 TCTCTCTTGAGCCGCCCTTCCCTCTGCTGCTCCCTCCAGATGGGGGTTAGTCCATAT 645
Db 1628 TAAGCTTTTCCCTTCTCTTACACCTGCGCATGCTCCCAATTTGGCAATTAATTCATCT 1687
QY 646 TTAACATGCCCATCCCTCAACAGTTGCTTAACATGCAAGTGTCTTTCTTTAGGACCCCTTC 705
Db 1688 TTAACATGCTGTTCTGTTAGTGTGCTAACTTAAGTGTCTTTCTTTATAGAACCCCTTC 1747
QY 706 TTAACGAGCAATATGCTGACCTGTACTATAAGATCTTCTGATAATGATTCGAGATT 765
Db 1748 TGACTGAGCAATATGCTTCTTATGATTAATAATCTTCTGATAATGCAATTAAGATT 1806
QY 766 TTTTCTGATAGTAGTAGAAGTGTCTGTTTCTTCACTTCCCTTTACTACGTGACTAGT 825
Db 1807 TTTTCTGATTAGTAAGAGTGTCTTCCATGTTAC-----TTTATTTCAGAGCTAATAAGT 1861
QY 826 GCTTCCCTGCTTTTCTAGTAACTGGTGTAGAAATCAAGTGTGCGGCTTTACAGTTT 885
Db 1862 GCTTTCCTTAGTTTCTAGTAACTAGTGTGTAATAATCATGTGTGACGCTTTATAGTTT 1921
QY 886 TAAACTATTATTAGATA-----TTCTGAAACATCACTGTTCCACAG 928
Db 1922 TAAATATTATTAGTAATTTCTTAACTATGAACCTTCTTAACTACTGTTTCCAGAT 1981
QY 929 TACCAACACTGTCTGATGATGATGCGGCCCTCTAGACCTCACCCACCGGACACATG 988
Db 1982 TACCGACACTGTCTGATGATGATGCGGCCCTCTAGACCTCACCCACCGGACACATG 2040
QY 989 CTTCCCGGTA 997
Db 2041 CTTCCCTGTA 2049

RESULT 12

AK054993
LOCUS
DEFINITION
Homo sapiens CDNA FLJ30431 fis, clone BRACE2008968, highly similar to RAS-RELATED C3 BOTULINUM TOXIN SUBSTRATE 1.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Unpublished
2 (bases 1 to 2315)
Isogai, T., Otsuki, T. and Sugiyama, T.
Direct Submission
Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan: cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB.

HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.

```

FEATURES
  source
    Location/Qualifiers
      i..2315
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="BRACE2008968"
      /tissue_type="cerebellum"
      /clone_lib="BRACE2"
      /note="cloning vector: pWE185FL3"
      603 a      531 c      501 g      680 t
      BASE COUNT
      ORIGIN

```

Query Match 28.6%; Score 290.6; DB 9; Length 2315;

Best Local Similarity 64.5%; Pred. No. 4.8e-75;
Matches 664; Conservative 0; Mismatches 304; Indels 61;

QY	1	CCCGTATCTTGTCTAGATTAAAGATTGCGCAAAATACCTTGTGAACTAAGTTGC--GTT	57
Db	1078	CCCCCATCTTGTGTTCAGATTAAAGAGTTGCCAAAATACCTTCTGAACTACACACTGCAATGTT	1137
QY	58	GTGCTGAGAACACGTAAGACTAAGCTGTTGCAGAGACTTTGTCTCTTAAGAGACTGCAGC	117
Db	1138	GTCCGAGAACACCGGACACTGAACCTTTGCAAGACCTTCGTCTTTGAGAAGACGGTAGC	1197
QY	118	TTCTGGGCTCAGG--GGTGCAGACCCCTCCCGTAGC-----TCCGACACCGTGTGACACA	169
Db	1198	TTCTGCAGTTAGGAGGTGCAGACACTGTCTCTCTATGTAGTTCTCAGATGCGTAAGACA	1257
QY	170	GCACAGCTCCTTAATGACACGGTGCCATGTAAACGACCTGTAACTTATCAGGCCCATGCT	229
Db	1258	GAACAGCCTCCGGAATGAAGCGTTGCCAATGAACTCACCACTGAGTTAGCAGACAGTGTT	1317
QY	230	CATTACTGAACCTTTGACTGTACGTACGATCGGTGTAAACGCTCTGCTCTTTGATTTCA	289
Db	1318	CCGACATAACATTTGACTGTAAATGGAGTGAGCGGTAGCAGCTCAGCTCTTTGGATCAGTC	1377
QY	290	TAGTGAAGTTCTCTAAATAATACAGCTGACCGGCTTCTGCAGGCTTTGAACAGAACTCTGGC	349
Db	1378	TTTGTGATTTTCATGCGAGTTTTCTGCACAGCTTTTGGCGAGATTTTGAACAGAACTG--	1435
QY	350	TCCTGTGTTGCCCTTAACGAAGTATCTGTTCCTAGTCGTGGGTGTCTCGGTGGAGTGT	409
Db	1436	-----CTATTTCTCTAATGAAGAAATCTGTGTT--TAGCTGTGGGTGTGCCGGTGGGGTGT	1489
QY	410	GTGAACACAGACCTCATCAAGAGAGACAGACAGTATTTTGAC--TAATATGAAGTAGAGAT	468
Db	1490	GT-----GTGATCAAGAGACAAAGACAGTATTTTGCACAAAATACGAAGTGGAG--	1537
QY	469	TAATTTACACTACATTTGATCATGCGAGTAA--TTCAACTCAATAAAAGTGTCCAGGGTAA	525
Db	1538	--ATTACACTACATGTACAAGGAATGAAGTGTACGGGTAAAAACTCTAAAGGTTA	1595
QY	526	AGCTTTTAAAGGTTAATTTCTGTCAACAGTAGATGACAAATGGCGGATCTTTATCAGTG	585
Db	1596	ATTCTGTCAATGCGAGTAGATGATGAAGAAAGGTTGGTATTATCAGGAAATCTTTCT	1655
QY	586	TCTCTCTTGAGCCCGCTTCCCGCTGCTGTCCCTCCCGAGATGGGGGTTTGAATCCATAT	645
Db	1656	TAAAGCTTTTCTCTTCTCTTACGCTGCATGCGCTCCCAAAATTTGGGCATTTAATTCATCT	1715
QY	646	TTAACTGGCGCATCTTCACAGTTGCTAACTTAGCAGTGCCTTTCTTTTAGGCCCCCTTC	705
Db	1716	TTAAACTGGGTTGTTCTGTAGTCGTTAACTTAGTAGTGTCTTTCTTATAGAACCCCTTC	1775
QY	706	TTAACGAGCAATATGTCTGACCTGTACTATAAGATCTTTCTGATAATGCATTCGGAGATT	765
Db	1776	TGACTGAGCAATATGCCT--CCTTGTATTATAAATCTTCTGATAATGCATTTAGAAGTT	1834
QY	766	TTTTTGGTAGATAGTAGAAGTGGGTTCCCTGTTTTCACCTTCCTTTACTCAGCTGACTAGT	825

Db	1835	TTTTGTGCGATTAGTAAAGTGCTTTCCATGTTAC-----TTTATTGAGAGCTAATAAGT	1885
Qy	826	GCATCCCTTCGGTTTTCTAGTAACTGGGTGTAGAAATCACGTGCTGGGCTTTTACAGTTTT	885
Db	1890	GCATCCCTTCGGTTTTCTAGTAACTGGGTGTAGAAATCACGTGCTGGGCTTTTACAGTTTT	1949
Qy	886	TAAACTATTTTATAGATA-----TTCTGAACATCATCTGCTTGGCCAGAG	928
Db	1950	TAAATATTTTATAGATAATCTTAAACTATGAACTTCTTAAACATCATCTGCTTGGCCAGAT	2009
Qy	929	TACCAACACTGTCATGTGATGATGCGGCCCTCTAGACCTCACCACCGGACACATG	988
Db	2010	TACCGACACTGTCACCTTGACCAATACTG-ACCCCTCTTTTACCTCGCCACCGGACACACG	2068
Qy	989	CTTCCGGTA 997	
Db	2069	CTTCCGTGA 2077	
RESULT 13			
LOCUS	HSAL132695	28567 bp	DNA linear PRI 09-NOV-2000
DEFINITION	Homo sapiens rai1 gene.		
ACCESSION	AJ132695		
VERSION	AJ132695.5 GI:8574037		
KEYWORDS	alternative splicing; Alu; AluJ; AluSg1; AluSg; AluX; AT-rich; C-rich; MIR; rai1 gene; rai1 protein; rai1b protein; repetitive sequence.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 Matos, P., Skaug, J., Marques, B., Beck, S., Verissimo, P., Gespach, C., Boavida, M.G., Scherer, S.W. and Jordan, P.		
TITLE	Small GTPase RAI1: structure, localization, and expression of the human gene		
JOURNAL	Biochem. Biophys. Res. Commun. 277 (3), 741-751 (2000)		
MEDLINE	20517245		
PUBMED	11062023		
REFERENCE	2. (bases 1 to 28567)		
AUTHORS	Jordan, P.		
TITLE	Direct Submission		
JOURNAL	Submitted (03-FEB-1999), Jordan P., Centro de Genetica Humana, Laboratorio de Oncobiologia, Instituto Nacional de Saude, Dr. Ricardo Jorge, Avenida Padre Cruz, 1649-016 Lisboa, PORTUGAL		
COMMENT	On Jun 20, 2000 this sequence version replaced gi:7248282.		
FEATURES	Related sequence AI638561.		
source	Location/Qualifiers		
	1..28567		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
	/chromosome="7"		
	/map="7p22"		
	/clone="H. NH0425A05, RPCI 11"		
repeat_region	1..109		
	/rpt_family="AluSg"		
repeat_region	125..422		
	/rpt_family="AluSc"		
repeat_region	441..742		
	/rpt_family="AluSg"		
gene	743..2817		
	/gene="rai1"		
promoter	743..1226		
	/gene="rai1"		
repeat_region	911..983		
	/rpt_family="GC-rich"		
misc_feature	936		
	/gene="rai1"		
misc_feature	1026		
	/notes="alternative transcription start site"		
misc_feature	1026		
	/gene="rai1"		

```

misc_feature
/note="alternative transcription start site"
1037
/gene="rac1"
/note="alternative transcription start site"
1053
/gene="rac1"
/note="alternative transcription start site"
1106..1175
/rpt_family="GC rich"
Join(1227..1261,13701..13772,18413..18530,22804..22860,
24268..24330,26010..26169,26458..26588)
/gene="rac1"
/note="alternative"
/codon_start=1
/product="Rac1b protein"
/protein_id="CAA10733.6"
/db_xref="GI:8574039"
/db_xref="SPTREMBL:O95501"
/translation="MQAICVVVGAGVGKTCLLSYTTNAPFGYIPTVFNYSANV
MVDGKPNVLGLWDTAGEDYDLRLPLSPQTVGETYGKDIISRGKDKPIADYELICFS
LYSPASFENVRAKWPEVRHHCNPPIILVGHKLDLDDKDIIEKLKKKLLPTIPYFQ
GUAMAKEIGAVYLECSALTQRLKTVFDEAIRAVLCPPPVKKRKKCLLL"
Join(1227..1261,13701..13772,18413..18530,24268..24330,
26010..26169,26458..26588)
/gene="rac1"
/codon_start=1
/product="Rac1 protein"
/protein_id="CAB53579.5"
/db_xref="GI:8574038"
/translation="MQAICVVVGAGVGKTCLLSYTTNAPFGYIPTVFNYSANV
MVDGKPNVLGLWDTAGEDYDLRLPLSPQTVGETYGKDIISRGKDKPIADYELICFS
HHCNPPIILVGHKLDLDDKDIIEKLKKKLLPTIPYFQGLAMAKEIGAVYLECSAL
TORGLKTVFDEAIRAVLCPPPVKKRKKCLLL"
<1227..1261
/gene="rac1"
/number=1
1262..13700
/gene="rac1"
/number=1
1390..1419
/rpt_family="GC rich"
1752..1793
/rpt_family="FLAM"
1794..2005
/rpt_family="AluJo"
2140..2171
/rpt_family="Simple_repeat"
2656..2972
/rpt_family="AluSx"
2980..3267
/rpt_family="AluSx"
3268..33289
/rpt_family="(T)n"
3802..4102
/rpt_family="AluJb"
4122..4386
/rpt_family="AluSx"
4538..4719
/rpt_family="LTR33A"
4754..5053
/rpt_family="AluSx"
5297..5594
/rpt_family="AluSx"
5655..5757
/rpt_family="AluSx"
5758..6062
/rpt_family="AluSx"
6063..6198
/rpt_family="AluSx"
6222..6401
/rpt_family="AluSg/x"
7162..7238
/rpt_family="L1MC5"
repeat_region
/rpt_family="Simple_repeat"
7348..7367
/rpt_family="AluSx"
7368..7656
/rpt_family="AluSx"
7880..8183
/rpt_family="AluSx"
8484..8778
/rpt_family="AluSx"
9071..9365
/rpt_family="AluSx"
10279..10583
/rpt_family="AluSp"
10609..10864
/rpt_family="L2"
10970..11276
/rpt_family="AluJo"
11276..11594
/rpt_family="AluJb"
11744..11825
/rpt_family="L2"
11764..11987
/rpt_family="MIR"
12989..13307
/rpt_family="AluSx"
13701..13772
/gene="rac1"
/number=2
13773..18412
/gene="rac1"
/number=2
13954..14235
/rpt_family="AluSx"
14807..14831
/rpt_family="Simple_repeat"
14836..15128
/rpt_family="AluSx"
15408..15715
/rpt_family="AluSx"
15872..15896
/rpt_family="AT rich"
15902..16186
/rpt_family="AluSx"
16187..16314
/rpt_family="AluJo"
16413..16585
/rpt_family="Simple_repeat"
16586..16679
/rpt_family="FLAM_A"
16707..16810
/rpt_family="L1M4"
16811..17108
/rpt_family="AluSg"
17109..17345
/rpt_family="L1M4"
17377..17686
/rpt_family="AluSx"
17703..17729
/rpt_family="Simple_repeat"
18413..18530
/gene="rac1"
/number=3
18531..22803
/gene="rac1"
/number=3
19287..19436
/rpt_family="MIR"
21161..21184
/rpt_family="Simple_repeat"
21185..21351
/rpt_family="AluSp"
21365..21403
/rpt_family="Alu"
21364..21365
misc_difference

```



```
/gene="rac1"
/nt="2345 nt insertion sequence in RPII 11-derived BAC

Query Match      28.6%; Score 290.6; DB 9; Length 28567;
Best Local Similarity 64.5%; Pred. No. 5.9e-75;
Matches 664; Conservative 0; Mismatches 304; Indels 61; Gaps 13;

QY 1 CCCCTATCTTCTGCTCAGATTGAAGATTGCCAATACCTTGTGACTAAGTTGC---GTT 57
Db 1 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 26871 CCCCAATCTTGTTCAGATTGAAGATTGCCAATACCTTGTGACTAAGTTGC---GTT 26930
Db 26871 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 58 GTGCTGAGAACACGCTAAGCTGTTGAGAGACTTTCCTTAAAGAGACTGACG 117
Db 58 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 26931 GTGCGGAGAACACGAGCACTGAATTTGCAAGACCTTCCTTGTGAAGAGCGTAGC 26990
Db 26931 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 118 TTCTGGGCTCAGG-GGTGCGACGCTCCCGTAGC-----TCCGAGCCGTTGACACA 169
Db 118 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 26991 TTCTGCGATTAGGAGGTGACACACTTCTCTCTATGTAGTTCTCAGATCGTAAAGCA 27050
Db 26991 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 170 GCACAGCTCTTAAATGACGCTGCCATGTAAAGCAGCTGTAATCATGACGCCATGCT 229
Db 170 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 27051 GAAAGAGCTCCCGAATGAAGCGTTGCCAATGAATCAACAGTAGTGTAGCAGACGTTT 27110
Db 27051 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 230 CATTACGTAATCTTGTACTGACGTACGATGGGTGAACAGCTCTGCTCTTTGATTCA 289
Db 230 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 27111 CCCGACATAACATTTGACTGTAAAGGATGAGCGTAGCAGCTCAGCTTTTGGATCAGTC 27170
Db 27111 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 290 TAGTGAGTTCTTAAATACAGCTGACCGGCTTCTGAGGCTTTGACAGAACTCTGGC 349
Db 290 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 27171 TTTGTGATTTTCATAGCGAGTTTCTGACAGCTTTTGGGAGATTTTGAACAGAACTG-- 27228
Db 27171 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 350 TCCTGTGCTGCTTACAGAGATTTCTGTTCTAGTCTGGGTGCTGGGTGAGTGT 409
Db 350 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 27229 ---CTATTTCTCTAATGAAGATTTCTGTT--TAGCTGTGGGTGTGCGGGTGGGGTGT 27282
Db 27229 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 410 GTGAACACAGCGTCTCAACAGGAGACAGACAGTATTTTGAC-TAATATGAAGTAGAGAT 468
Db 410 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 27283 GT-----GTGATCAAGGACAAAGACAGATTTTGAACAAATACGAAGTGGAG-- 27330
Db 27283 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 469 TAATTTACACTACATTTGACATGAGTAA---TTCAACTGAATAAAAGTGTACGGGTAA 525
Db 469 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 27331 --ATTTACACTACATTTGACAGGAATGAAGTGTACGGGTAAATAAACTCTAAAGGTTA 27388
Db 27331 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 526 AGCTTTTAAAGGTTAATTTCTGTCACAGTAGATGACAAATGCGGATCTTATCAGTG 585
Db 526 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 27389 ATTTCTGTCAAATGCGATGATGATGAAGAAAGAGGTTGGTATATCAGGAATGTTTCT 27448
Db 27389 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 586 TCTCTCTTGAGCCCCCTTCCCTGCTGCTCCCTCCCGAGATGGGGTGGAGTCCATAT 645
Db 586 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 27449 TAAGCTTTTCTTCTTCTTACACTGCGATGCCCTCCCAAAATGGGATTTAATCATCT 27508
Db 27449 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 646 TTAACATGGCCATCCTCAGATGCTGCTAATAGTACAGTGTCTTTTCTTAGGACCCCTTC 705
Db 646 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 27509 TTAACATGGTGTCTGTTAGTCTAATAGTGTGCTTTCTTCTTATAGAACCCCTTC 27568
Db 27509 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 706 TTAACGAGCAATATGCTGACCTGCTACTATAGATCTTTCTGATAATGCAATGCGAGATT 765
Db 706 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 27569 TGACTGAGCAATATGCTC-CCTTGATATATAAAATCTTCTGATATGCAATGCAATGAGGTT 27627
Db 27569 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 766 TTTTGTGATAGTAGAAGTGGCTTCTGTTTTCACCTTCTCTTACTCAGCTCAGTAGT 825
Db 766 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 27628 TTTTGTGATAGTAGAAGTGGCTTCTTCCATGTTAC-----TTTATTCAGAGCTAATAGT 27682
Db 27628 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 826 GCTTCCCTTCGTTTCTAGTAATGCGGTGAGAAATCAGCTGCTGCGGCTTTACAGTTTT 885
Db 826 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 27683 GCCTTCCTTAGTTTCTAGTAATGAGTGTAAATCATGTTGTCAGCTTTATAGTTTT 27742
Db 27683 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 886 TAAACTATTTTAGATA-----TTCTGAACATCATCTGTTGCGCAGAG 928
Db 886 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 27743 TAAATAATTTAGATAATTTCTTAACTATGAACCTTCTTAACTACTGCTTGGCCAGAT 27802
Db 27743 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 929 TACCACACTGTATGTGATTGATGCGGCCCTCTAGACTCACCACCGGACACATG 988
Db 929 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
Db 27803 TACCAGACTGCTACTTGACCAATACTG-ACCCTTTTACCTGCGCCACGGGACACAG 27861
QY 989 CTTCCGGTA 997
Db 27862 CTTCTGTGA 27870

RESULT 14
AC009412 212827 bp DNA linear PRI 07-NOV-2001
LOCUS Homo sapiens BAC clone RP11-425P5 from 7, complete sequence.
AC009412
AC009412.6 GI:14190769
VERSION HTG.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 212827)
AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 212827)
AUTHORS Hou, S., Maupin, R., Haakenson, W., Gregory, S. and Belter, E.
TITLE The sequence of Homo sapiens BAC clone RP11-425P5
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 212827)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (21-AUG-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 212827)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (23-MAY-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 212827)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (07-NOV-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On May 23, 2001 this sequence version replaced gi:13431187.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
-----
Center project name: H_NH0425P05

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and
sequencing collaboration between the NHGRI Chromosome 7 Mapping
Project (Eric D. Green, Director), John D. McPherson in the
Department of Genetics (Washington University), and the Washington
```

University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send mailto:egreen@nhgri.nih.gov, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCL11 human BAC library was made from the blood of one male donor, as described by Ooegawa, K., Woon, P.Y., Zhao, B., Frenken, E., Tatenio, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is CTD-2195P2, 200 bp overlap. Actual start of this clone is at base position 1 of RP11-425P5; actual end is at base position 212633 of RP11-425P5.

FEATURES

source

Location/Qualifiers

1. 212827

/organism="Homo sapiens"

/mol_type="Genomic DNA"

/db_xref="taxon:9606"

/chromosome="7"

/map="7"

/clone="RP11-425P5"

/clone_lib="RPCL11"

1623..1696

/rpt_family="MIR"

1698..2006

/rpt_family="Alu"

2024..2337

/rpt_family="Alu"

3006..3093

/rpt_family="L1"

3453..3552

/rpt_family="L2"

4924..5046

/rpt_family="ERV1"

5318..5693

/note="similar to EST BF996284 (NID:G12402607)"

5580..5845

/rpt_family="Alu"

5925..6277

/rpt_family="MIR"

6842..7041

/note="similar to EST BF369507 (NID:G11331532)"

7364..7661

/rpt_family="Alu"

8029..8048

/rpt_family="Alu"

8379..8678

/rpt_family="Alu"

8560..8762

/note="similar to EST AA699398 (NID:G2702592) z140a03.s1"

8720..9031

/rpt_family="Alu"

9083..9560

/note="similar to EST AA633603 (NID:G2556817) ae66b02.s1"

9598..9899

/rpt_family="Alu"

9871..10098

/note="similar to EST AA699398 (NID:G2702592) z140a03.s1"

10746..11044

/rpt_family="L1"

11075..11101

/rpt_family="AT-rich"

11102..11372

/rpt_family="Alu"

11411..11465

/rpt_family="GA-rich"

11552..11644

repeat_region 11685..12193 /rpt_family="L2"

repeat_region 12194..12486 /rpt_family="L1"

repeat_region 12487..13003 /rpt_family="Alu"

repeat_region 13004..13305 /rpt_family="L1"

repeat_region 13306..13501 /rpt_family="Alu"

repeat_region 13503..13892 /rpt_family="L1"

repeat_region 13893..14224 /rpt_family="L1"

misc_feature 14244..14544 /note="similar to EST BE061102 (NID:G8405752)"

repeat_region 14546..14828 /rpt_family="Alu"

repeat_region 15091..15189 /rpt_family="Alu"

repeat_region 15672..16449 /rpt_family="MIR"

repeat_region 16640..16844 /rpt_family="Achobo"

repeat_region 16917..16959 /rpt_family="Alu"

repeat_region 16960..17287 /rpt_family="Achobo"

repeat_region 17288..17572 /rpt_family="Alu"

misc_feature 17610..17938 /rpt_family="Achobo"

repeat_region 17793..18068 /note="similar to EST BF996427 (NID:G12402750)"

repeat_region 18251..18537 /rpt_family="Alu"

misc_feature 18485..18894 /note="similar to EST BE062347 (NID:G8406997)"

misc_feature 18661..18809 /note="similar to EST BG186544 (NID:G13708231)"

repeat_region 18957..19160 /rpt_family="L1"

repeat_region 19314..19528 /rpt_family="L1"

repeat_region 19819..20135 /rpt_family="Alu"

repeat_region 21101..21269 /rpt_family="L1"

repeat_region 21882..22203 /rpt_family="L1"

misc_feature 22429..22797 /note="similar to EST BF993478 (NID:G12399801)"

repeat_region 22717..23007 /rpt_family="L1"

misc_feature 22993..23191 /note="similar to EST BF767044 (NID:G12115035)"

repeat_region 23167..23471 /rpt_family="Alu"

misc_feature 23511..23930 /note="similar to EST AW105611 (NID:G6076346) xd49g02.x1"

repeat_region 25387..25677 /rpt_family="Alu"

Query Match 28.8%; Score 290.6; DB 9; Length 212827;
Best Local Similarity 84.5%; Pred. No. 7e-75;
Matches 664; Conservative 0; Mismatches 304; Indels 61; Gaps 13;

QY 1 CCCTATTCTTGTTCAGATTGAAGATTGCCAAATACCTTGTGAACCTAAGTTGC--GTT 57

Db 174899 CCCCATTCTTGTTCAGATTGAAGATTGCCAAATACCTTGTGAACCTAAGTTGTT 174958

QY 58 GTGCTGAGACACGTAAAGCACTAGCTGTTGAGAGACTTTGCTCCTTAAGAGACTGCAGC 117

174959 GTGCCGAGAACCCAGCAGTCTGACTTTGCAAGACCTTCCTCTTGGAGAGACGTAGC 175018
118 TTCTGGCTCAGG-GGTGCGACACCTCCCGTAGC-----TCCAGACCGGTGTGACACA 169
175019 TTCTGCGATTAGAGGTGCAGACACTTCTCTCTATGTAGTCTCAGATCGTAAGACA 175078
170 GCACAGCTCTTAAATGACAGCTGCCATGTAACGCACTGTAACTATACAGCCATGCT 229
175079 GAACAGCTCTCCGAATGAAGCGTTGCCAATGAATCAACAGTGTAGTACAGCAGCTGT 175138
230 CAATACCTAATCTTGTACTGTACGTACAGTGGGTGTAACAGCTCTGCTCTTTGATTCA 289
175139 CCCGACATAAATTTGATGTAATGAGTGAAGTACAGTCTAGCTCTTTGGATCAGTC 175198
290 TAGTGAGTCTCTAAAATACAGCTGACGCGCTTCTGAGCGCTTTGACAGAACTCTGGC 349
175199 TTTGTGATTTTCATAGCGAGTTTCTGACCGAGTTTTCGCGAGATTTTGAACAGAACTG-- 175256
350 TCCTGTCTGCTCTAAGAGTATCTGTTCTAGTCTGGGTGCTGGGTGAGTGT 409
175257 -----CTATTTCTCTAATGAAGATTCGTT--TAGCTGTGGGTGTCGCCGGTGGGTGT 175310
410 GTGAAACACGACGTCATCAAGGAGACAGACAGTATTTGAC-TAATATGAAGTAGAGT 468
175311 GT-----GTGATCAAGGACAAAGACAGTATTTGACAAATACGAAGTGGAG-- 175358
469 TAATTTACACTATGTTGATGATGAGTAA--TTCAACTGAATAAAAGTGTACGGGTAA 525
175359 --ATTACACTACATTTGATCAAGGATGAAGTGTACGGGTAAATACTCTAAAAGTTA 175416
526 AGCTTTTAAAGGTTAATTTCTGTCACACAGTACATGACAAATGCGCGATCTTACAGT 585
175417 ATTTCTGCTCAATGCTAGATGATGAAGAGGTTGGTATTAATCAGGAAATGTTTCT 175476
586 TCTCTCTTGAGCCCTTCCCTGCTGTCCTCCCGACAGATGGGGTGTGAGTCCATAT 645
175477 TAAGCTTTTCTTCTTCTACACCTGCACTGCTCCCGAAATGGGCATTTAATCACT 175536
646 TTAACATGGCCATCTCAGATGCTAACTTAGCAAGTGTCTTTTATAGAACCCCTTC 705
175537 TTAACATGGTGTCTGTTAGTCTAACTTAGTAAAGTGTCTTTTATAGAACCCCTTC 175596
706 TTAACGACCAATATGCTGACCTGTACTATAAGTATCTTCTGATAATGCAATTCGAGAT 765
175597 TGACTGACCAATAGCTC-CTTGATTTATAAATCTTCTGATAATGCAATTTAGAGGT 175655
766 TTTTGTGTAGATAGTAGAAGTGGCTTCTGTTTTCACCTTCTTACTCAGCTGACTAGT 825
175656 TTTTGTGCGATTAGTAAAGTGTCTTCCATCTTAC-----TTTATTCAGAGCTAATAGT 175710
826 GCTTCCCTCGTTTCTAGTAACTGGGTGTAGAAATCAGGTGTGCGGCTTTACAGTTT 885
175711 GCTTTCCTTAGTTTCTAGTAACTAGGTGTAAATAATCATGTGTGACGCTTTATAGTTT 175770
886 TAAACTATTTTAGATA-----TTCTGAAACATCACTGCTTCCGAG 928
175771 TAAATATTTTAGATAATCTTAACTATGAACCTTCTTAACTCACTGCTTCTGCCAGAT 175830
929 TACCACACCTGTCTATGTAATGATGCGCCCTCTAGACCTCACCCACGCGGACACATG 988
175831 TACCGACACTGTCACTTGACCAATCTG-ACCTCTTTTACCTGCCACCGGACACACG 175889
989 CTTCGGGTA 997
175890 CTTCTCTGTA 175898

RESULT 15
AX408985
LOCUS 1232 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence from Patent WO0229103.
ACCESSION AX408985

AX408985.1 GI:21441690
Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE
Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.
Gene expression profiles in liver cancer
Patent: WO 0229103-A 1632 11-APR-2002;
GENE LOGIC INC (US)
FEATURES
Location/Qualifiers
1..1232
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/note="EMBL/GenBank Accession No. D25274"
BASE COUNT 329 a 240 c 251 g 411 t 1 others
ORIGIN
Query Match 28.0%; Score 285.2; DB 6; Length 1232;
Best Local Similarity 64.4%; Pred. No. 1.9e-73;
Matches 658; Conservative 0; Mismatches 303; Indels 61; Gaps 13;
QY 8 TCTTGCTCAGATTAAAGATTGCCAAATATCTTGTGAACTAAGTTGC---GTTGTGTGA 64
Db 1 TCTTGTTTCAATTAAAGATTGCCAAATATCTTGTGAACTAAGTTGC---GTTGTGTGA 60
QY 65 GAACACCTAAGCATAAGCTGTTGAGAGACTTGTCTTAAAGAGCTGACAGCTCTCTGGG 124
Db 61 GAACACCGAGCACTGAATCTTAGAAAGACCTTGTCTTTGAGAAGCGGTACTGTGCA 120
QY 125 CTCAGG-GGTGCGACACCTCCCGTAGC-----TCCCAGACCCGTGTGACACAGCAGC 176
Db 121 GTTAGGAGGTGACAGACCTTGTCTCTCTATGTAGTCTCAGATCGTAAAGCAGACAGC 180
QY 177 CTCTTAAATGACAGCTGCCATGTAACGACCTGTGTAACTTATCAGCCCATGCTCATACG 236
Db 181 CTCGGAATGAAGCGTTGCCAATTAAGTCACTCAGTGTAGTGTAGCAGCAGCTTCCCGACA 240
QY 237 TAATTTGTACTGTACGTACAGTGGGTGTAACAGCTCTGCTCTTGTATTCATAGTACG 296
Db 241 TAACATTTGTACTGTATGAGTGTAGCGTAGAGCTCAGCTCTTTGGATCTAGTCTTTGTA 300
QY 297 TTCTCTAAAATACAGCTGACCGGCTTCTGACGCTTTTGAACAGAACTCTGGCTCTGTG 356
Db 301 TTTCTATAGCGAGTTTCTGACCAAGCTTTTCGGAGATTTTGAACAGAACTG-----CTA 354
QY 357 TTGCTCTTAAGAGTATTTCTGTTCTCTAGTGTGGGTGTGCTGGTGGAGTGTGTGAAC 416
Db 355 TTTCTCTTAATGAAGATTTCTGTT--TAGCTGTGGGTGTGCGGGTGGGTGTGT----- 407
QY 417 ACGAGCTCAATCAAGGAGACAGACAGTATTTTGAC-TAATATGAAGTAGAGATTAATTA 475
Db 408 -----GTGATCAAGGACAAAGACAGTATTTTGACAAATACGAAGTGGAG----ATT 458
QY 476 CACTACATTTGATACAGAGT---AATTTCACTGAATAAAGTGTCTACGGGTAAAGCTTTT 532
Db 459 CACTACATTTGATCAAGGATGAAGTGTACGGGTAAAGTCTTAAAGCTTTAATTTCTG 518
QY 533 TAACGGTTAATTTCTGTCAACAGTAGTACAAATGGCCGATCTTATCAGTGTCTCTCT 592
Db 519 TCAATATGCTAGTAGATGAAGAAAGGTTGTTATTTATCAGAAATGTTTCTTAAAGCTT 578
QY 593 TGAGCCCTCTCCCTCTGCTGCTCCCTCCAGATGGCGGTGAGTCCATATTAAGT 652
Db 579 TTCTCTTCTTACACTGCTGCTGCTCCCAATTTGGGCAATTAATCATCTTTAAACT 638
QY 653 GGCCATCTCTCACAGTGTCTAACTTAGCAAGTCTTTTCTTTAGACCCCTCTCTTAACGA 712
Db 639 GGTGTTTCTGTAGTGTGCTAAGTGTAGTGTCTTTTCTTATAGAACCCCTCTCTGACTGA 698
QY 713 GCAATATGCTGACCTGTACTATAAGATCTTCTGTGATTAATGCAATTCGAGATTTTGTG 772

[illegible]

Search completed: November 23, 2003, 14:02:01
Job time : 3817.98 secs